



0500
#2

169.1770

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:)
JAMES ROBERT METCALFE ET AL.) : Examiner: Not Yet Assigned
Application No.: 09/615,781) : Group Art Unit: NYA
Filed: July 13, 2000) :
For: AUTOMATED DOCUMENT) :
PRODUCTION FROM A) :
SEARCH ENVIRONMENT) : July 31, 2000

Commissioner for Patents
Washington, D.C. 20231

CLAIM TO PRIORITY

Sir:

Applicants hereby claim priority under the
International Convention and all rights to which they are
entitled under 35 U.S.C. § 119 based upon the following
Australian Priority Application:

PQ1638, filed July 14, 1999.

A certified copy of the priority document is
enclosed.

This Page Blank (uspto)

Applicants' undersigned attorney may be reached in our New York office by telephone at (212) 218-2100. All correspondence should continue to be directed to our address given below.

Respectfully submitted,


Attorney for Applicants

Registration No. 29,296

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200

NY_MAIN 99396 v 1

This Page Blank (uspto)

09/615781



Patent Office
Canberra

I, ANNA MAIJA EVERETT, ACTING TEAM-LEADER EXAMINATION SUPPORT & SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 1638 for a patent by CANON KABUSHIKI KAISHA filed on 14 July 1999.

I further certify that pursuant to the provisions of Section 38(1) of the Patents Act 1990 a complete specification was filed on 12 July 2000 and it is an associated application to Provisional Application No. PQ 1638 and has been allocated No. 47166/00.



WITNESS my hand this
Eighteenth day of July 2000

A. M. Everett.

ANNA MAIJA EVERETT
ACTING TEAM LEADER
EXAMINATION SUPPORT & SALES

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

This Page Blank (uspto)

ORIGINAL

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION FOR THE INVENTION ENTITLED:

Automated Document Production from a Search Environment

**Name and Address
of Applicant:**

Canon Kabushiki Kaisha, incorporated in Japan, of 30-2,
Shimomaruko 3-chome, Ohta-ku, Tokyo, 146, JAPAN

Names of Inventors: James Robert Metcalfe and John Stewart Reeves

This invention is best described in the following statement:

AUTOMATED DOCUMENT PRODUCTION FROM A SEARCH ENVIRONMENT

Technical Field of the Invention

The present invention relates generally to network-based browsing and searching applications and, in particular, to the preparation of printable documents that facilitate the rapid and discriminating review of search results obtained from such applications, and for providing advertising and like messages.

Background Art

In order to facilitate the accessing of information available through computer networks such as the Internet and the World Wide Web ("the Web"), network service providers typically allow users access to one or more search engines that are operable by the user to identify specific classes of information available on the network.

Fig. 1 depicts an example of the Web browser application and a typical presentation via a graphical user interface (GUI) of a search engine page 100 that may arise from a user searching of a particular keyword string of text 102, in this example STRING. STRING may be a single alphanumeric word, or a list of such words, perhaps linked by Boolean operators and represents the search criteria used by the search engine. As illustrated in Fig. 1, search results 104 are typically ranked according to the quality of hit upon the searched string and typically lists the location 106 of the search result and its title 108. The actual presentation of the search results 104 is often user definable within a range of settings established by the search engine. Typically a location 106 of an individual result 114 is expressed as a Uniform Resource Location (URL). In some cases, a corresponding title 108 and the location 106 are combined as a Uniform Resource Indicator (URI). Often, the search engine application provides for some amount of text 110, typically a few lines or an abstract, relating to the particular location to be

presented as part of the search result. In many search engines, this text typically represents the first few lines of text of the referenced location.

As a consequence, based on the presentation shown in Fig. 1, the user is then able to scroll through the search results 104 using a scroll bar 112 or by selecting different display pages of the search result to identify those individual results desired to be reviewed. In those circumstances where the user preselects the text 110 to be displayed, often the search results 104 are provided arranged ten to a page, however removal of the text 110 can provide the search results to be provided at a rate of approximately twenty per page, where "page" in this fashion represents electronic page of information displayed on a display screen to the computer network user.

Many problems exist with the above described arrangement. Firstly, where only the URL 106 is presented in the result, often the user has no means of interpreting the search result other than by accessing the URL. Where the titles 108 are selected for display, often the particular title 108 provides no information as to the specific content, or context of that content, to be found at the corresponding URL, or bearing any relationship to the searched string 102. Further, where the text 110 is also provided, there is no guarantee that the searched string 102 will be presented in the text displayed on the search page 100. Further, in any configuration there is no guarantee that when the user accesses the particular location 106, at which the searched string 102 is purported to be found, that the searched string will actually be found. As a consequence of the inadequacies of the information presented in the search page 100 shown in Fig. 1, users often spend excessive amounts of time accessing individual locations, reviewing the locations and, where appropriate, discarding the relevance of that location before referring to a further location.

Fig. 2A exemplifies a display of a Web page 200 accessed through a user selecting the first search result 114 shown in Fig. 1. As seen from Fig. 2A, the displayed

Web page 200 includes a title banner 202, various images 204, 206, 208, 210, and text 212 incorporating the searched string 214. With such an example, accessing of the first search result 114 by the user provides an immediate result in response to the user's searching of the search string.

5 However, in Fig. 2B, a Web title page 230 for another search result 116 shown in Fig. 1 is shown which also includes a title 216, a certain amount of text 218, an animated GIF image 220 together with a number of further URL links 222 within the same primary location depicted by the particular URL, in this case URL#m. Notably, the search string is not seen in the displayed page 230 of Fig. 2B. In order for the user to find the search
10 string, it is necessary for the user to scroll through the Web page 230 using the scroll bar 112. As seen from Fig. 2C, the search string 102 is located at 232 within a display screen 234 within the Web page 230 defined by the URL#m. The display screen 234 is seen to have images 224, 226 and text 228.

Further, where the user reviews information at any one location, the only
15 convenient way of forming a reasonable record of that review is to print a particular page of the reference location. Print facilities provided with browser applications and search engine pages are limited to one Web page at a time. This requires the user to access each Web page and to then print that page where appropriate.

It will therefore be appreciated that traditional methods of viewing search results
20 obtained over computer networks can be time consuming, and are not conducive to providing a convenient record of search results.

Disclosure of the Invention

It is an object of the present invention to substantially overcome, or at least ameliorate, one or more disadvantages of existing arrangements.

25 In accordance with one aspect of the present invention there is disclosed a method of presenting search results obtained from a search conducted over a computer

network, said search including searching criteria and returning information including a plurality of network locations, said method comprising the steps of:

- (a) extracting data from a first said first network location;
- (b) examining said data to identify therein said searching criteria to provide
5 at least one specific location within said first network location of said searching criteria;
- (c) using said one specific location to identify from said extracted data specific data including at least said searching criteria;
- (d) formatting said specific data into a printable document; and
- (e) repeating steps (a) to (d) for each remaining said network location in
10 which step (d) appends said formatted data of said remaining network location to said printable document.

In accordance with another aspect of the present invention there is disclosed a method of formatting an electronic document intended for reproduction by printing, said method comprising the steps of:

- 15 (a) sourcing main data from at least one location in a computer network, said data including a plurality of data types;
- (b) formatting said data into a common data type suitable for each of electronic display and printing;
- (c) arranging said formatted data as a printable document spanning at least
20 one printable page;
- (d) identifying one or more locations where said at least one page is void of said formatted data; and
- (e) sourcing further data configured in said common type and sized to be positioned within said one or more locations; and
- 25 (f) formatting said further data within said one or more locations in said printable document.

In accordance with another aspect of the present invention there is disclosed a method of formatting an electronic document intended for reproduction by printing, said method comprising the steps of:

- 30 (a) obtaining from a searching process location information within a computer network of at least one search result returned by said searching process;
- (b) using said location information to fetch data from said computer network relating to each said search result, said data including said searching criteria; and

(c) formatting the fetched data including said searching criteria into a printable electronic document.

Apparatus for performing each of the methods are also disclosed.

Brief Description of the Drawings

5 A number of preferred embodiments of the present invention will now be described with reference to Appendix A and the drawings, in which:

Fig. 1 shows an exemplary view of a typical search engine page as seen using a Web browser application;

10 Figs. 2A to 2C depict example pages that may be found through traditional examination of the search engine page of Fig. 1;

Fig. 3 is a schematic exemplary representation of a search result page according to the preferred embodiment based upon the search result of Figs. 2A to 2C;

Fig. 4 is a flow chart depicting results collection and formatting according to the preferred embodiment;

15 Fig. 5 is a schematic illustration of a network traversal page according to another embodiment;

Fig. 6 is a flow diagram depicting generation of the arrangement shown in Fig. 5; and

20 Fig. 7 is a schematic block diagram representation of a computer system in which the preferred embodiments may be implemented.

Detailed Description including Best Mode

To assist users in being able to track and trace their traversal of computer networks such as the Web, Canon Information Systems Research Australia Pty Ltd has developed a "Hypertext Document Collating Tool" which is currently the subject of
25 United States Patent Application No. 08/903,743 filed 31 July 1997 (Attorney Ref: 378728US CFP0568US Page+20), the disclosure of which is annexed hereto as

Appendix A. The Hypertext Document Collating Tool operates in a background mode behind the browsing software application used to traverse the Web so as to automatically and transparently create a printable document that includes the various Web sites and documents encountered by the user during the traversal of the Web. The Web sites and documents typically include numerous data types, components and configurations, such as simple ASCII text, JPEG images, GIF and TIF static and animated graphics, and so on. Such varied source data is often termed "hypertext" and is formatted primarily for electronic representation via a display screen, but not necessarily for hard copy reproduction. A preferred implementation of the disclosure of that patent application is realised by the product marketed under the trade mark WebRecord™ by Canon Kabushiki Kaisha and Canon Information Systems Research Australia Pty Ltd.

The embodiments described in the present specification are preferably implemented as additional features to the Hypertext Document Collating Tool of Appendix A. However, the present invention is not limited to user with the Hypertext Document Collating Tool, WebRecord™, or similar products, but has wider application and may for example be implemented in generic browsing software or searching arrangements, as will be appreciated by those skilled in the art having read and understood this specification.

The preferred embodiment provides an arrangement whereby the user, having conducted a search using traditional browsing software or search engines, such as described with reference to Fig. 1, is not required to individually examine each of the locations identified in the search result. According to the preferred embodiment, a further application, herein referred to as a formatted document generator application, operating in a manner akin to the aforementioned Hypertext Document Collating Tool, receives input information from the search engine result in the form of the searched keyword string and

5

10

25

contained in the search string 102. The portion 320 below the graphical divider 318 incorporates that portion of the Web page which includes the searched string 102 and thus provides the specific content relating to the string as it is found in that Web page. With respect to URL#1, this will be seen to be consistent with that substantially displayed in the arrangement of Fig. 2A.

This separation and presentation of information is more generically depicted in Fig. 3 with respect to URL#2 with respect to the top of the Web page context information being separated from the content information including the searched string.

The example shown in Fig. 3 with reference to URL#m illustrates the specific power of the preferred embodiment in avoiding a need of the user to review and/or print the entire contents within any one search result URL location. As seen, the section relating to URL#m includes a contextual title and page corresponding to that shown in Fig. 2B followed by a sectional divider and a section corresponding to the information contained in Fig. 2C. Significantly, use of the formatted document generator application permits passing over, or skipping, those portions of the referenced location (in this case URL#m) that don not contain the searched string. In some situations, this can obviate the necessity to scroll through or print a single Web page that may occupy many display screens of information, which may also occupy many individual sheets of paper in a hard copy reproduction. The preferred embodiment acts to condense the particular information found at the location identified by the search result into a convenient, interpretable and manageable form.

As seen, the printable document 302 depicted in Fig. 3 incorporates information from a number of Web locations identified by the search result, the information being presented in a contextual form and also in a form in which the particular content relating to the search string is identified.

As seen in Fig. 3, each of the portions 320 includes a reproduction 324 of the actual URL of that portion thereby maintaining a record of the location of the referenced information. The portions 322 may also be configured to identify a similar location 326, but in this case, modified by the relative location within the referenced URL as to where the searched string is found. for example, for URL#1, the searched string is found on "page 1", representing the electronic screen page number corresponding to that seen in Fig. 2A. For URL#m, the page number is "page n", representing the number of electronic screens required to be conventionally scrolled by the user within URL#m in order to locate the searched string.

A permanent record may be obtained either by saving the electronically displayed printable document 302 to memory, for example by actuating a SAVE icon 324, or by printing the electronically displayed printable document, for example by actuating a PRINT icon 326.

According to the preferred embodiment, the printable document of Fig. 3 can be generated once the search engine results of Fig. 1 are returned to the user via the browser application. Generation of the printable page of Fig. 3 occurs as a result of the hypertext document collating and formatting, and without additional burden upon the Web browser application or search engine, by directly and separately accessing the individual URLs returned by the search engine result and searching within each URL for the relevant searched string information. This method is depicted in the flow chart of Fig. 4.

Fig. 4 depicts the operation 400 of a Web browser 402 and associated search engine 404 and their relationship to the operation of the formatted document generator application 430 according to the preferred embodiment. As seen, the browser 402 incorporates the search engine 404 which acts upon a search string 406 entered by the user. The search string 406 may be one or more words of text separated or associated

using Boolean operators. The search engine 406 returns a search result 408, typically being an individual page displayed by the browser 402 and which traditionally replicates the searched string together with an indication of the total number of results 410 or hits identified by the search. The user is then able to select groups 414 of results typically based upon a ranked number which can be displayed on a page at any one time. As seen, results 412 incorporating Results # 1-m are displayed.

At this stage, the user may invoke the operation of the formatted document generator application 430 to create a printable document based upon the search result 412 indicated in the search page. If this is not desired, operation of the application 430 is not performed and the user is free to continue utilising the search engine 404 or browser 402 in a traditional fashion, for example by effecting a further search or directly examining any one of the search results 412. Hypertext document collating and formatting according to Appendix A may be performed if such is pursued.

Alternatively, and according to the preferred embodiment, where the user elects for the creation of a printable search result in step 432, step 434 checks that the printable document is to be formed for the displayed results, in this case, Results # 1-m. Where appropriate, the user may select, via an interconnection 422 to the search result 408, for another group 414 of the search results 412 to be selected. Once the group of results is settled in step 434, step 436 copies search information data including the searched string 406 via an interconnection 418, and the URL's corresponding to the selected group of results 412 via an interconnection 420. The interconnections 416, 418, 420 and 422 shown in Fig. 4 will be appreciated as being illustrative of interactions between the applications 402 and 430, and those skilled in the art will appreciate that such may be implemented in various forms or procedures, not necessarily dependent on the individual "interconnections" as shown, any one of which may be optionally implemented. For

example, interconnections 418 and 420 may be unitarily formed as a single request for search information, such as when the search information is derived from a file, the file for example being formed by batch processing of search engine calls.

In step 438 which follows, the generator application 430, independently of the
5 browser 402 or search engine 406, fetches the data from the Web at the various URL's
given by the individual results 412. Step 440 then commences a processing loop on the
fetched/downloaded data, one result at a time. At this stage, determination is made in
step 440 as to whether or not there are any unprocessed results and, where there exist no
unprocessed results, step 442 follows. In step 442, examination is made of the fetched
10 data of the particular result 412 to identify if the searched string 406 is actually found
therein. In those cases where the string being searched is not found within the Web site, a
situation which occurs all too frequently and much to the chagrin of Internet users,
step 442 returns control to step 440 so as to process the next result. Where the searched
string is found in step 442, step 444 follows which records the specific location of the
15 search string 406 within the particular Web site.

Processing of the individual result location is then performed to format the
search result. Next, in step 446, the top of the Web page for the result being processed is
extracted from the fetched URL data. In step 448 this is converted from a hypertext
format to a common format suitable for both electronic display and hard copy printing,
20 and incorporated into a printable document. Step 450 which follows inserts a result
divider, corresponding to the graphic 318 of Fig. 3, into the formatted document.
Step 452 then uses the recorded location from step 444 to extract a particular section of
the Web page that incorporates the searched string. In step 454 which follows, the
extracted selection incorporating the searched string is formatted into the common
25 displayable and printable form and incorporated into the printable document. This is

followed by step 456 which inserts the location break, corresponding to the separator 316 of Fig. 3, into the printable document. Control then returns to step 440 for the processing of the next result. Where there are no further results to be processed, step 440 transfers control to step 458 which enables the user to view the printable document via a display
5 screen in the fashion shown in Fig. 3, from which search results may be saved and/or printed by actuating an appropriate icon.

In an alternative implementation, steps 438, 442 and 444 may be combined into a single process where, as the data is fetched and downloaded from the Web, the data is simultaneously checked to identify the searched string, and where the searched string is
10 identified, that specific location within the data is recorded. Where the searched string is not located in the URL accessed data, that data may be discarded, without further processing, so that data from the next URL in the search results list can be fetched.

According to the preferred embodiment, the formatted document generator application extracts only predetermined portions of each search result URL thus ensuring
15 that the printable search result document 302 of Fig. 3 does not contain extraneous matter not necessarily related to the search. This avoids consuming inordinate computing resources for generation and reproduction of the printable document.

Once reproduced, by display or printing, the user may closely review the search result presented by the printable document 302 and thereafter access particular URLs that
20 may be desired for closer examination and review. A traditionally formatted printable document may then be created for any closely examined URL, according to the principles described in Appendix A.

Where desired, the formatting performed in step 454 of the extracted section incorporating the searched string may be made in such a manner so as to highlight the
25 particular search string as it appears with in the formatted document. In this fashion, any

person reviewing the printable document of Fig. 3 either in its electronic or hard copy printed form, has their attention directed to the particular search string as it is reproduced.

It will be further appreciated that the purpose of the search is to identify the searched string, and thus the display of the extracted top of page for the result is not essential for the performing of the present invention since this is not necessarily important to the particular search result. However, the present inventors consider that the incorporation that such of page information is relevant so that the actual searched string as found is placed in some user interpretable context. For example, a search of the term "automobiles" may return results relating to the manufacturers of automobiles configured for use on the road. However, the same search may return a result for a manufacturer of toy or model automobiles suitable for the playing of children's games and the like. In many instances, the names of such automobiles and trade marks associated with such automobiles may be the same irrespective of whether they are real automobiles or toy automobiles. The incorporation of the contextual top of page information will generally assist the user in distinguishing between toy motor vehicles and real motor vehicles.

The printable document generated according to the preferred embodiment may be arranged to extend over one or more printable pages which can be viewed via the display shown in Fig. 3 using the vertical scroll bar 308. Since the GUI display of Fig. 3 is a "print preview" type display, it has a fixed width and a horizontal scroll bar is not necessary. Further to the user's selectability of the display of the top part of each Web page, the preferred embodiment may be configured for the display of more than one section which incorporates the searched string, each such section for example being separated by replication of the graphic divider 318. For example, preferences may be established for reproducing a predetermined number of hits upon the search string and/or all of the hits on the search string.

Further, the amount of information reproduced relevant to any one hit on the searched string, and as a consequence of step 452, may be varied according to user preferences. By default, such may include a single displayable screen taken from a Web page incorporating the searched string. Alternatively, the entirety of the Web page at a particular location may be referenced. In a further alternative, where the text string forms part of a paragraph of text, only that paragraph or section of text may be extracted for reproduction. In a further user selectable optimisation, because many images found on the Web are not directly reproducible in print, examples of which being animated GIF images and the like, the user may select disablement of such images in the printable document thereby permitting the formatted document generator application to reconfigure the printable document compared to the actual Web produced location so as to optimise the amount of text to be reproduced. Other image types, such as JPEG and static GIF images may be disabled from printing as desired. This may be important, particularly when handling "home pages", which represent a root directory URL. In such pages, and many others, user selectable icons and the like often consume much of the displayable page but often provide no substantive information, particularly in satisfying a search query. Appropriate configuration of the formatted document generator, can permit such icons and like objects to be excluded from the printable document, thus affording greater levels of compaction of information relevant to the searched string in the printable document.

As a consequence, the printable document 302 of Figs. 3 and method of Fig. 4, permit network obtained search results to be automatically formatted and collated into a single, user interpretable, document that provides for compaction of more than one, and typically numerous, summaries of each search result on any single printable page. The embodiment also provides a permanent record of the search and search results in a

convenient and compact form. Significantly, the information is presented to the user in the same manner that such would be seen had the user actually accessed the referenced URL in the traditional fashion, thus ensuring maintenance of the context of the referenced URL and the searched string to the user. This is to be contrasted with prior art search result presentations such as Fig. 1, which are not configured to reproduce graphics, images and other indicia that provide user interpretable context to the search result. Prior art search result presentations typically carry little or no information of the searched string.

A further embodiment is illustrated in Fig. 5 where a further printable page is illustrated as part of a GUI incorporating sections configured according to the previous embodiment.

As will be appreciated by those skilled in the art, automated collating and formatting of the printable documents of Figs. 3 and 5 can result in the content of any printable column or page extending to a further column or page thus causing the printable document to be somewhat non-contiguous. Whilst such may be often tolerated, where the document is segmented in the fashion shown in Figs. 3 and 5, it may be desirable for each individual segment to be formed in a single column upon a single page. Such can result in the creation of vacant spaces at the end of columns into which no specific information may be placed due to size limitations.

According to the further embodiment of Fig. 5, such vacant spaces may be occupied through the insertion of a printable message such as an advertisement. The message or advertisement may be inserted by a server associated with the generation of the printable document and/or the information contained therein. In a preferred implementation, as seen in Fig. 5 where a printable document 500 generated is formed from search results according to the embodiment of Fig. 3, the printable document 500

includes a page break 502 separating a first printable page 504 from a second printable page 506, and a first column 508 into which are neatly placed two search result sections 510 and 512. However, in this example, the search result returns only one further result 514 which does not occupy the entirety of a second column 516 of the page 504, leaving spaces 518 and 520 respectively at the top and bottom of the column 516. According to the present embodiment, advertisements are sourced based upon the subject matter being searched, in a manner consistent with traditional electronic advertisements that can occur during Web browsing sessions, the sourced advertisements being formatted into a printable form and inserted into the spaces 518 and 520. However, unlike such traditional Internet advertisements, which are configured for transient display upon an electronic display (eg. video screen), advertisements of the present embodiment are intended ultimately for reproduction by means of printing and, as a consequence, are specifically configured for such display and will therefore, for example, be absent animated graphical objects or other moving components. The data format of such message is therefore preferably common to that used in sections 510 and 512.

Using the foregoing example, where the user searches the string "automobile", an advertisement 522 may be placed in the space 518 advertising "AUSSIE Motor Vehicles" in the manner illustrated in Fig. 5. Another advertisement 524, for example relating to automobile parts such as those manufactured by "JIM's Mag Wheels", may be inserted into the otherwise vacant space 520.

According to the present embodiment, one or more of the printable messages or advertisements may be returned to the formatted document generator in a number of ways. Firstly, this may be by way of the particular search engine being used to conduct the search on the string which traditionally returns electronic messages for display of the video screen of the user. Any such returned advertisement may then be interpreted by the

formatted document generator which then converts into a printable form suitable for formatting and placement into the printable document of Fig. 5. This method, and another, may be described with reference to Fig. 6.

Fig. 6 shows a system 600 in which a user 602 operates a Web browser application 604 together with a formatted document generator application 606. The user 602 accesses a search engine server 610 via a network 612 in order to identify information available via the network 612. The search engine 610 is associated with an advertisement server 614 which includes a keyword algorithm interpreter 616 for examining the user's search string to thereby return an appropriate electronically displayable advertisement 622 via the search engine server 610 to the user 602. The formatted document generator 606 then interprets the advertisement and formats the advertisement into a form suitable for electronic printing by depositing the same into an electronic printable document 608, which, as will be observed, corresponds to the representation of Fig. 5.

In an alternative, the formatted document generator 606, having identified the need for the placement of an advertisement to fill the vacant column space, is configured to communicate via the network 612, without interaction with the browser 604 or search engine server 610, with a dedicated print advertisement server 618 for the provision of a printable advertisement to the formatted document generator 606. The print advertisement server 618 is configured in a fashion to examine the network 612, for example including the advertisement server 614 and/or itself, to identify one or more printable advertisements 620 to be automatically returned, preferably via the server 618, to the formatted document generator 606. In this fashion, rather than relying upon the search engine server 610 for the presentation of advertisements, the formatted document generator 606 may call the dedicated server 618 to provide advertisements specifically

configured for reproduction by both display and printing. In this fashion, the formatted document generator 606 either directly, or via the server 618, may control those advertisements that may be placed into the otherwise blank spaces 518 or 520 of the document 500. Such an arrangement, where calls are made via the server 618, permit
5 monitoring of advertisements returned for printing in the document 608.

The specific advantage of incorporating the printable advertisements in the printable document is that the printable document can become a permanent record of the user's search of the Web that may be required for later review. As a consequence, the advertisement that is printed with the search result also becomes a permanent record of
10 the advertisement and thus can be interpreted as substantially more valuable in an advertising sense than a transient advertisement, such as those traditionally displayed on electronic display apparatus via the search engine server 610 as will be known to those skilled in the art. As a consequence, the individual printable advertisements can be provided at a premium cost compared to those transient advertisements. Further, where
15 the formatted document generator 606 interacts with the server 618 for accessing the printable advertisements 620, the dedicated server 618 may be configured for direct management of costs associated with advertising charges associated with providing the printable advertisement 620 to the user 602 for incorporation into the printable document 608.

Formatted document generation described with reference to Figs. 3 to 6 is preferably practiced using a conventional general-purpose computer system 700, such as that shown in Fig. 7 wherein the processes of Figs. 4 and 6 and the Hypertext Document Collating Tool may be implemented as software, such as an application program executing within the computer system 700. In particular, the steps of the methods of
25 Figs. 3 to 6 are effected by instructions in the software that are carried out by the

computer. The software may be divided into two separate parts; one part for carrying out the methods, and another part to manage the user interface between the latter and the user. The software may be stored in a computer readable medium, including the storage devices described below, for example. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable medium having such software or computer program recorded on it is a computer program product. The use of the computer program product in the computer preferably effects an advantageous apparatus for formatted document generation in accordance with the embodiments of the invention.

The computer system 700 comprises a computer module 701, input devices such as a keyboard 702 and mouse 703, output devices including a printer 715 and a display device 714. The display 714 is used to reproduce the GUI and images depicted in Figs. 1, 2A-2C, 3 and 5, whilst the printer 715 may be used to print the printable documents 300 and 500 of Figs. 3 and 5 respectively. A Modulator-Demodulator (Modem) transceiver device 716 is used by the computer module 701 for communicating to and from a communications network 720, for example connectable via a telephone line 721 or other functional medium. The modem 716 can be used to obtain access to the Internet, and other network systems, such as a Local Area Network (LAN) or a Wide Area Network (WAN).

The computer module 701 typically includes at least one processor unit 705, a memory unit 706, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 707, and an I/O interface 713 for the keyboard 702 and mouse 703 and optionally a joystick (not illustrated), and an interface 708 for the modem 716. A storage device 709 is provided and typically includes a hard disk drive 710 and a floppy disk

drive 711. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 712 is typically provided as a non-volatile source of data. The components 705 to 713 of the computer module 701, typically communicate via an interconnected bus 704 and in a manner which results in a conventional mode of operation of the computer system 700 known to those in the relevant art. Examples of computers on which the
5 embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

Typically, the application program of the described embodiments is resident on the hard disk drive 710 and read and controlled in its execution by the processor 705.
10 Intermediate storage of the program and any data fetched from the network 720 may be accomplished using the semiconductor memory 706, possibly in concert with the hard disk drive 710. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 712 or 711, or alternatively may be read by the user from the network 720 via the modem device 716.
15 Still further, the software can also be loaded into the computer system 700 from other computer readable medium including magnetic tape, a ROM or integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 701 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including e-mail transmissions and information recorded on
20 websites and the like. The foregoing is merely exemplary of relevant computer readable mediums. Other computer readable mediums may be practiced without departing from the scope and spirit of the invention.

The methods of the described embodiments may alternatively be implemented in dedicated hardware such as one or more integrated circuits performing one or more
25 functions or sub-functions of the formatted document generator. Such dedicated

hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories.

Industrial Applicability

Embodiments of the present invention are applicable to network data accessing
5 and retrieval systems and the described embodiments are intended to complement existing
browsing and searching tools, particularly in Internet and World Wide Web applications.
It is also noted that the embodiment of Figs. 5 and 6 is not limited to for use with the
embodiment of Figs. 3 and 4, but finds general application for use with the Hypertext
Document Collating Tool described in Appendix A, and also in like arrangements.
10 Further, whilst the embodiment of Figs. 3 and 4 describes an arrangement that is adjunct
to the search engine application, an alternative embodiment may be integrated into the
search engine application so as to automatically supplement or replace, the provision of
search results as depicted in Fig. 1. For example, the search engine may be configured to
generate a file which incorporates the search string, the URL's returned in the search
15 result, and user preferences for the formatted document generation. Such preferences
may include the number of URL's to be examined for any one search result (eg. the first
ten), the amount of information to be extracted from any referenced URL (eg. the top of
page, the number of hits on the searched string, the size of the extracted string containing
portion), and a maximum size (eg. in terms of printable pages) of the printable document
20 for any one searched string, to name but a few. The provision of such a file, and
appropriate configuration of the formatted document generator application to interpret the
file, can facilitate the batch processing of search strings and corresponding document
generation without user interaction.

Further, with respect to the embodiments of Figs. 5 and 6, rather than being arranged within columns, the advertisements 522 and/or 524 may be arranged to extend across the width of the page, for example like a "banner" style advertisement.

Also, whilst the embodiment of Figs. 3 and 4 has been described with reference
5 to Internet and World Wide Web searching, the inventive concept is not limited thereto, but applies generally to all computer networks. For example, a Local Area Network within an office environment may incorporate many thousands of word processing documents distributed amongst many computer devices. The described embodiments may thus be used to perform keyword searching on such documents in order to identify
10 specific classes of documents without the user having to open each individual document.

The foregoing describes only one embodiment/some embodiments of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the invention, the embodiment(s) being illustrative and not restrictive.

15 In the context of this specification, the word "comprising" means "including principally but not necessarily solely" or "having" or "including" and not "consisting only of". Variations of the word comprising, such as "comprise" and "comprises" have corresponding meanings.

CLAIMS:

1. A method of presenting search results obtained from a search conducted over a computer network, said search including searching criteria and returning information including a plurality of network locations, said method comprising the steps of:
 - (b) extracting data from a first said first network location;
 - (b) examining said data to identify therein said searching criteria to provide at least one specific location within said first network location of said searching criteria;
 - (c) using said one specific location to identify from said extracted data specific data including at least said searching criteria;
 - (d) formatting said specific data into a printable document; and
 - (e) repeating steps (a) to (d) for each remaining said network location in which step (d) appends said formatted data of said remaining network location to said printable document.
- 1A. A method according to claim 1 wherein step (c) identifies further specific data from a predetermined plurality of said specific locations within said network location.
2. A method according to claim 1 wherein step (d) further comprises formatting root data, obtained from said extracted data at a root location of said first network location, and said specific data into said printable document whereby said root data supports a contextual basis to said specific data.
3. A method according to claim 1 wherein (first) graphical separators are incorporated into said formatted printable document to separate extracted data from different ones of said network locations.
4. A method according to claim 2 wherein (second) graphical separators are incorporated into said formatted printable document to separate said specific data from said root data of one said network location.
5. A method according to claim 1 wherein said printable document is formatted into a plurality of columns.

6. A method according to claim 1 comprising the further step of:

(f) electronically displaying the printable document to an instigator of said search.

5 7. A method according to claim 6 wherein said printable document is displayed in a print preview format, known per sé.

8. A method according to claim 1 or 6 comprising the further step of:

(g) printing said printable document.

10

9. A method according to claim 1 wherein said extracted data is of a plurality of data types, and said formatting converts said data types into a common data type suitable for each of electronic display and printing.

15 10. A method of formatting an electronic document intended for reproduction by printing, said method comprising the steps of:

(a) sourcing main data from at least one location in a computer network, said data including a plurality of data types;

20 (b) formatting said data into a common data type suitable for each of electronic display and printing;

(c) arranging said formatted data as a printable document spanning at least one printable page;

(d) identifying one or more locations where said at least one page is void of said formatted data; and

25 (e) sourcing further data configured in said common type and sized to be positioned within said one or more locations; and

(f) formatting said further data within said one or more locations in said printable document.

30 11. A method according to claim 10 wherein step (c) comprises arranging at least some of said formatted data so that plural components thereof are reproducible on at least a single said printable page.

12. A method according to step (e) comprises sourcing said further data configured in a plurality of data types and step (f) comprises formatting said further data into said common data type.

5 13. A method according to claim 10 wherein said further data is related to a context of said main data.

14. A method according to claim 13 wherein said further data comprises advertising content related to said context.

10 15. A method according to claim 13 wherein step (e) comprises advising a specified location within said network of said context, and said specified location then using said context to extract from a further location with said network said further data for supply to said printable document.

15 16. A method of formatting an electronic document intended for reproduction by printing, said method comprising the steps of:

(a) obtaining from a searching process location information within a computer network of at least one search result returned by said searching process;

20 (b) using said location information to fetch data from said computer network relating to each said search result, said data including said searching criteria; and

(c) formatting the fetched data including said searching criteria into a printable electronic document.

25 17. A method according to claim 16 wherein step (b) comprises, for each said search result, the sub-steps of:

(ba) using the corresponding location information to fetch all data accessible for said search result;

(bb) checking said fetched all data to identify said searching criteria therein;

30 and

(bc) recording, for each identification of said searching criteria, a specific location of said searching criteria within said fetched all data.

18. A method according to claim 17 wherein step (b) comprises, for each said search result, the further sequential step of:

(bd) discarding said fetched all data where said searching criteria is not at least once identified therein.

19. A method according to claim 16, 17 or 18 wherein step (c) comprises, for each said search result, the sub-steps of:

(ca) identifying at least one portion of said fetched data associated with at least one occurrence of said search criteria;

(cb) converting said identified one portion to a data format suitable for both electronic display and printing; and

(cc) incorporating said converted identified one portion into said electronic document according to a predetermined format.

20. A method according to claim 19 when dependent on claim 17 wherein, for each said search result, step (ca) utilises said specific location to identify the corresponding said portion.

21. A method according to claim 19 wherein step (c) comprises, for each said search result, the further sub-steps of:

(cd) identifying an initial portion of said fetched data arranged at a root of said location information; and

(ce) converting said initial portion to said data format;

wherein step (cc) comprises incorporating said converted initial portion and said converted identified one portion into said electronic document.

22. A method of formatting a printable document substantially as described herein with reference to Figs. 3 and 4,

23. A method of formatting a printable document substantially as described herein with reference to Figs. 5 and 6 of the drawings.

24. A printable document formed according to the method of any one of the proceeding claims.

25. Apparatus configured to form a printable document according to claim 24.

Dated this Fourteenth Day of July 1999

CANON KABUSHIKI KAISHA

Patent Attorneys for the Applicant

Spruson & Ferguson

HYPER-TEXT DOCUMENT FORMATTING COLLATING AND PRINTING

Field of Invention

The present invention relates to hyper-text documents and, in particular, to the network access, formatting and printing of hyper-text documents.

5 Background of the Invention

Many computer based document mark-up languages have been developed in order to allow computer-aided document preparation. Examples of such languages include TROFF, TeX, RTF, as well as many proprietary formats associated with computer hosted word processing applications. These mark-up languages are designed to allow
10 the computer assisted preparation of a document destined for printing. As a consequence to these developments, the prevalence and active nature of digital computers has encouraged the introduction of hyper-links in documents.

A hyper-link is a pointer, typically embedded in a document, that provides a direct link to another portion of the same document, another document, another
15 resource, available on the current network node or another network node. Hyper-links are often used on the Internet, and in particular the World Wide Web to link a document at one Web site with a document at another Web site.

Hyper-links are only operational when a document is viewed on-line, and not when the document is in printed form. The increased value of these on-line hyper-text
20 documents has caused a weakening of the previous focus on printing. New generation languages used to interpret hyper-text linked documents such as SGML and HTML (Hyper-Text Mark-up Language), have few features to support the description of their printed form. More importantly, because the principle value of hyper-text documents is for on-line viewing, these documents are formatted by their authors in a manner which
25 is appropriate for screen viewing, and not necessarily for viewing in printed form.

As a result it is now the case that very large quantities of information are recorded in network accessed on-line services in formats which are appropriate for screen based viewing, but not as appropriate for viewing in printed form. Further,

because printing is not a focus of applications which access these hyper-text documents (that is, hyper-text browser applications), their printing facilities are generally poor.

Common problems encountered when printing hyper-text documents include:

- information is broken up into small hyper-text documents, and many documents
5 need to be collated to form a desired body of information;
- text is formatted with fewer words per line than is common for printed pages, and
in general the density of information is less than is typical for printed pages;
- hyper-text document viewing programs are document-centric, that is they operate
on a single hyper-text document at a time, which results in this being the unit of
10 printing, resulting in much repetitive work by the user to print a set of linked hyper-text
documents, and typically no more than one hyper-text document on each printed page;
- hyper-text document viewing programs generally do not print all the features of
hyper-text pages which are displayed on-screen (a display device), in particular the
target of hyper-links is often not included in printouts.

15 It is possible for the provider of a hyper-text document designed for screen
viewing to also provide substantially the same document in a different form designed
for printing, but this requires double handling by the document provider. It also often
results in significant differences between the screen version of the document and the
printed form.

20 The problem of no more than one hyper-text document per printed page can
sometimes be addressed by the reduction and rotation of the image of each basic page
and printing each reduced page image on, say, one half of a printed page. However
this method does not save paper at a given scale. For example, if a large number of
small hyper-text documents are printed, each of which only occupies 25% of a printed
25 (physical) page, even though the documents are photo-reduced and printed two per
physical page, each physical page still has 75% blank space. Further, this method does
not provide continuous page-length columns. Continuous column printing provides
improved readability and space utilization.

An object of the invention is to substantially overcome at least one of the aforementioned problems in the formatting of hyper-text documents.

Summary of the Invention

In accordance with one aspect of the present invention there is provided a method
5 of collating hyper-text documents comprising the steps of:

- (a) monitoring a user's access patterns to said hyper-text documents;
- (b) accessing said hyper-text documents including structure information of the accessed hyper-text documents;
- (c) creating a formatted version of the accessed hyper-text documents for said
10 user.

In accordance with another aspect of the present invention there is provided a method of collating hyper-text documents comprising steps of:

- (a) accessing said hyper-text documents including structure information;
- (b) creating a formatted version of said accessed hyper-text documents wherein
15 said formatted version is characterised by a single or multiple column printing such that each printed page contains as many of said hyper-text documents as can reasonably fit in an available space on a printed page.

Other aspects and features of the present invention are also disclosed.

Brief Description of the Drawings

20 A preferred embodiment of the present invention will now be described with reference to the accompanying drawings in which:

Fig. 1 is a block diagram showing the operating environment of the preferred embodiment of the present invention;

Fig. 2 shows the visual appearance of a user interface in accordance with the
25 preferred embodiment.

Fig. 3 is a block diagram of an internal structure of the preferred embodiment of the invention;

Fig. 4 is a block diagram of a general purpose computer upon which the preferred embodiment of the present invention can be practiced;

Fig. 5 is an example of the display screen during hyper-text document preparation; and

Fig. 6 is a flowchart depicting operation of a hyper-text document formatting portion of the preferred embodiment.

Description of the Preferred Embodiment

5 The preferred embodiment of the present invention is described as a computer application program hosted on the Windows™ operating system developed by Microsoft Corporation. However, those skilled in the art will recognise that the described embodiment may can be implemented on computer systems hosted by other operating
10 systems. For example, the preferred embodiment can be performed on computer systems running UNIX™, OS/2™, DOS™. The application program has a user interface which includes menu items and controls that respond to mouse and keyboard operations. The application program has the ability to transmit data to one or more printers either directly connected to a host computer or accessed over a network. The
15 application program also has the ability to transmit and receive data to a connected digital communications network (for example the "Internet").

A high-level block diagram is illustrated in Fig. 1 to provide an overview of the preferred embodiment. A Hyper-text browser 10 is provided to output to a display device 11 for viewing hyper-text documents. Typically, the hyper-text browser 10 is of
20 the form of application software implemented on a general purpose computer system (eg. IBM PC or compatible, Apple Macintosh, Sun-Workstation etc.) and hyper-text documents include images, linked documents and simple TEXT documents. Current examples of the hyper-text browser include Microsoft Explorer and NETSCAPE The computer system (not shown in Fig. 1) usually forms an interface which connects a
25 network system 12 of computers to the display device 11 and to a print output device 13.

A hyper-text document formatter 14, preferably implemented as a software module on the general purpose computer, is operable to format a hyper-text document and controlled in part by instructions derived 15 from the hyper-text browser 10

responding to a user's request. Further, the hyper-text document formatter 14 communicates with the network system 12 to perform a multitude of functions including gathering, formatting, and collating documents with direct instructions from the hyper-text browser 10 or the user.

5 Referring to Fig. 2, there is shown a user interface layout of the preferred embodiment as displayed on the display device 11 and which comprises a menu and control area 21, a print list display 22, and a print preview display 23. The print list display 22 includes a list of print items 22A, 22B, 22C, each of which include a print item mark box 24, a hyper-text document title text field 25, a fetch status text field 26
10 and a location text field 27. The print list display 22 and the print preview display 23 are scrollable by means of scroll bar controls 28 and 29.

The print preview display 23 displays (shows) representations of the printed pages which are to be produced on the printer output device 13 using current selected print options, for example in a WYSIWYG ("what you see is what you get") format. The
15 user is free to select from the menu and controls 21 a print option other than the current print option. Such print option can include print settings for the print output device 13, portrait or landscape orientation of pages, print resolution and scaling. Upon user selection of an option, the current print preview display 23 is appropriately updated. However the display in the print preview display 23 is regenerated automatically as a
20 current application state changes without intervention required by the user. Application states which can effect the print preview display 23 include, but are not limited to, the currently selected printer, the currently selected paper type, formatting options which can be set by the operator, the set of marked items in a print list (ie. those selected by a mark in the print item mark box 24) and the order of marked items associated with the
25 print list.

The preferred embodiment of the invention can be practised using a conventional general-purpose (host) computer system, such as the computer system 40 shown in Fig. 4, wherein the application program discussed above and to be described with reference to the other drawings is implemented as software executed on the computer

system 40. The computer system 40 comprises a computer module 41, input devices such as a keyboard 42 and mouse 43, output devices including a printer 13 and a display device 11. A Modulator-Demodulator (Modem) transceiver device 52 is used by the computer module 41 for communicating to and from a computer network, for example connectable via a telephone line or other functional medium. The modem 52 can be used to obtain access to the Internet, and other network systems.

The computer module 41 typically includes at least one processor unit 45, a memory unit 46, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 47, and an I/O interface 48 for the keyboard 42 a mouse 43 and optionally a joystick (not illustrated). A storage device 49 is provided and typically includes a hard disk drive 53 and a floppy disk drive 54. A CD-ROM drive 55 is typically provided as a non-volatile source of data. The components 45 to 49 and 53 to 55 of the computer module 41, typically communicate via an interconnected bus 50 and in a manner which results in a conventional mode of operation of the computer system 40 known to those in the relevant art. Examples of computers on which the embodiments can be practised include IBM-PC/ATs and compatibles, Sun Sparcstations or alike computer systems. Typically, the application program of the preferred embodiment is resident on a hard disk drive 53 and read and controlled using the processor 45. Intermediate storage of the program and the print list and any data fetched from the network may be accomplished using the semiconductor memory 46, possibly in concert with the hard disk drive 53. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk, or alternatively could be read by the user from the network via the modem device 52.

Fig. 3 shows a block diagram representation of an internal structure of the preferred embodiment, which comprises a user interface task 30, a monitoring task 31, a data fetching task 32, a formatting task 33, an internal print list storage 34, the print list display 22 (also shown in Fig. 2), the print preview display 23, a temporary file storage 35, a network and file system interface 36, and a printer interface 37.

The internal print list storage 34 is structured as a list of records in the memory 46 of the general purpose computer system 40, each record being referred to hereinafter as a "print item". Each print item represents at least one hyper-text document, and comprises a Uniform Resource Locator (URL) by which the associated hyper-text document can be retrieved as well as a further list of records, each of which is referred to herein as a sub-item. Each sub-item represents a distinct file-like unit of data which is required to complete the formatting and displaying of the hyper-text document associated with the print item. These units of data (or sub-items) are most commonly hyper-text documents in HTML format and images in GIF or JPEG format.

10 Each sub-item records a file name within the temporary file storage where the unit of data will be, or is, stored.

In Fig. 3, the four tasks 30, 31, 32, 33 are shown, each of which is implemented as a separate thread within a single application process. The internal print list storage 34 is shared by the tasks 30-33 in a manner to avoid conflicts. Each task 30-33 gains access to the print list on the internal storage 34 by first obtaining a "mutex" lock (mutually exclusive lock). Once the lock is obtained, the task reads and possibly modifies the print list and then releases the lock. Upon release of the lock, if changes were made to the print list, messages are forwarded to the user interface task 30, the formatting task 33 and the data fetching task 32 to inform them that changes have been made.

15

20

The user interface task 30 performs user interface operations by having a waiting state 30A and by acceptance of user interface events such as clicks and movements of the mouse 43, responds to process 30B as appropriate to each event. Operation of the task 30 is achieved by a message loop structure processing each operating system generated event in turn and is linked to the print list display 22.

25

The monitoring task 31 performs monitoring 31A of user initiated access to documents including hyper-text documents using the hyper-text browser 10, and entering 31B each such document accessed by the user into the print list. In particular, the browser 10 includes an application program interface (API) which allows viewing

of information being cached by the browser 10. In this manner, the monitoring task 31 is able to take and maintain a record of the operation, typically sequential, of the browser 10. From the record, the print list 34 is automatically created using the URL's of the items located. The user is then able to edit the print list 34 by deselecting those
5 items not required to be printed.

The fetching task 32 performs fetching of all documents which are listed in the print list along with associated data necessary for producing a visually pleasing (desired) or viewable formatted version of the documents in print form. Typically, the associated data includes print settings for a print devices to which the documents are to
10 be directed. Operation of the fetching task 32 is preferably achieved through use of Internet protocols and/or network access techniques provided by the host operating system and includes a wait stage 32A for detecting any change in the print list, and a fetching stage 32B, for fetching the required data and storing the data in a temporary file storage 35 typically formed within the memory 46. The fetching task 32 is also
15 responsible for initiating further fetches and amending the print list accordingly. Amending the print list or adding to the print list hyper-text pages which are hyper-linked from one of the pages previously fetched, by the fetching task 32, is typically performed as a background task to the hyper-text browser 10. Hyper-links previously visited by the fetching task 32 are preferably not re-visited to avoid repetition. The
20 user may elect, as part of optional settings that the fetching task 32 visits, a predetermined number of hyper-link pages for augmenting the print list accordingly.

Preferably, the fetching task 32 provides a cross-referencing feature, should the user select or desire such option, which maintains a cross referencing to URL or hyper-links of hyper-text documents to be printed (formatted version) with an indexing of
25 cross references and a corresponding page (number) in the document to be printed.

In this connection, the formatted version includes a table of contents listing each hyper-text document represented in the document to be printed. Each entry in the table of contents is labelled with the position (page number) at which the associated hyper-text document occurs within the said formatted version.

The formatting task 33 performs formatting of all documents which are listed in the print list in a manner suitable for printed output, and also optionally showing a preview of the printed output which would be produced in the print preview area. Its operation is achieved by a recursive descent HTML parser and formatter, and results
5 from waiting 33A for a change in the print list, and a format stage 33B which formats the documents and forwards it to a printer interface 37 for hard copy reproduction.

Notwithstanding that the updating of the print preview display 23 appears, under some circumstances, to depend on an availability of a hyper-text document through the network, a substantial portion of the tasks described with reference to Fig. 3 are
10 performed substantially instantaneously in background mode unbeknown or at least not immediately apparent to the user. Typically, the tasks 30-33 can be performed synchronously or asynchronously with a user's access pattern. Usually, a user accesses or visits, with the aid of the browser application, root hyper-text documents. Described in an alternative way, hyper-text documents visited by a user are referred to herein as
15 root hyper-text documents, and any further hyper-links and their associated documents are visited and fetched by the fetching task 32 respectively. The depth to which hyper-links are followed in fetching hyper-text documents is user defined. Preferably, all hyper-links of a root hyper-text document having predetermined characteristics are visited by the fetching task 32 and the associated (hyper-text) documents are retrieved.
20 For example, a user may mark hyper-links to be followed to a predetermined depth or the user may specify characteristics of hyper-links, and their associated documents, to be all documents descendent from a current root hyper-text document containing predetermined keyword.

Fig. 5 provides an illustrative representation of the preferred embodiment use.
25 Fig. 5 shows a display screen 60 of the display 11 which has two windows clearly displayed. A window 70 is a web-browser application window that displays a text document 67 (corresponding to a few of the introductory paragraphs of this patent description). This forms a background window and is representative of the hyper-text browser application 10 covering the entire screen area. Superimposed on top of the

window 70 is a window 63 corresponding to a working display of the application program of the preferred embodiment, described earlier with reference to Fig. 2. The user in this case is preparing a document formed from three sources, each mentioned in the print display list 61. A first source 68, called FRED, is a simple text source
5 previously encountered during a Web review, and occupies a first position in the document being formed. A second source 69, being a picture of a vehicle, occupies a second position, whilst a third source, corresponding to the background text document 67, occupies the third position. It is seen from the print display list that a Search engine, used to locate the text document 67 has been de-selected (N-No) from
10 display, and hence does not appear in the WYSIWYG print preview 65. The display list indicates that each source has been fetched is its corresponding URL, and is selected (Y-Yes) for display. In each case the location identifier provides the Web site address for the source material.

As seen in Fig. 5, the second column 64 of the print preview 65 has a blank
15 section 66. As seen from the print display list 61, the text document 67 remains in a "fetching" state, where the text is being retrieved and formatted for WYSIWYG display. Once this is completed, the section 66 displays the text that has since been fetched and the print display list 61 is updated to indicate a "fetched" status for that document.

20 In compiling the print document, the application program, and in particular, the document formatter 33B, recognises that the width of FRED and the picture are narrower than the page, and therefore establishes a column corresponding to their width. Because of its length, the text document 67 is formatted, firstly into a narrower, left hand column 62 related to the width of FRED 68 and the picture 69, and then to
25 flow into the right hand column 64 which is adjusted to a width to substantially fill the page. Importantly, the application program is configured to automatically detect the selected content of a source, and to incorporate that content into the print preview display 23 (65) in an economical manner so that as many hyper-text documents as can reasonably be fitted to a page can be displayed. This reduces paper consumption.

The preferred embodiment is configured to operate in background mode whilst the user is traversing the World Wide Web to automatically create and format a printable document representing a chronological history of the user's traversal of the World Wide Web. Typically, the preferred embodiment operates in a background mode as a window operating behind a web browser window. As seen in Fig. 6, a flowchart of procedures 100 of the hyper-text document formatting portion of the preferred embodiment commences at a starting point 102. This entry point leads to a step 104 where the application attempts to read an HTML element from a Web document currently being viewed using a Web browser program. At step 106, which follows step 104, an assessment of data availability is made and if none is available, step 108 assesses whether or not another document can be opened. If so, control is returned to step 104 for handling the new document. If not, document formatting is completed at step 110.

If data is available at step 106, control is passed to step 112 where the HTML element of the current Web site location is formatted into a standard form able to be printed using the application program. At step 114, an assessment is undertaken as to whether or not the formatted element is able to fit on to the page to be printed. If so, control is transferred to step 118 where the formatted HTML document is emitted as an output document. If the formatted element does not fit on to the page as determined by step 114, control is passed to step 116 which splits off, or culls, the non-fitting remainder of the formatted element. This enables control to be passed to step 118 for emitting of the remaining formatted HTML document. After step 118, control is passed to step 120 which assesses whether or not there is a remainder, for example left over from step 116. If so, control is returned to step 112 so that the remainder can be formatted and processed in the manner described above. If there is no remainder, control is returned to step 104 in order to read the next HTML element.

With the arrangement described in Fig. 6, whilst the user browses the World Wide Web, the application program continually assesses the data being viewed in the browser window and automatically formats that data into a continuous printable

document displayed in the window for example shown in Figs. 2 and 5. When the user has completed browsing, the window of the application program (ie. window 63 of Fig. 5), can be selected. Using the print display list 61, the user can either select or deselect certain documents located during the Web browsing session for printing.

5 During the course of a browsing session, all documents seen are automatically enabled in the print document window. Accordingly, prior to printing all that is necessary is for the user to cull out or deselect those components not desired for printing. For example, if the user had made use of a search engine during the Web browsing session, there may be little point in printing out the text associated with that search engine. All

10 that would be necessary to print could be the actual document or Web site location found as a result of the search, such as shown in the example of Fig. 5.

A further advantage of the present invention is that, in the printed document, at the completion of each section relating to an individual Web location, the actual Web location is printed onto the printed document so that the user has a permanent hard

15 copy record of not only the information sourced but of the location of that source.

The foregoing only describes one embodiment of the present invention, however, modifications and/or changes can be made thereto by a person skilled in the art without departing from the scope of the invention.

WE CLAIM:

1. A method of collating hyper-text documents, said method comprising the steps of:
 - (a) monitoring a user's access patterns to said hyper-text documents; and
 - 5 (b) accessing said hyper-text documents including structure information of the accessed hyper-text documents;
 - (c) creating a formatted version of the accessed hyper-text documents for said user.
- 10 2. A method as claimed in claim 1, wherein steps (a), (b) and (c) are conducted while the user accesses hyper-text documents.
3. A method as claimed in claim 1, wherein said formatted version of the accessed hyper-text document is updated upon new hyper-text pages being accessed.
- 15 4. A method as claimed in claim 1, wherein said steps are performed in background mode.
5. A method as claimed in claim 1, wherein steps (b) and (c) are performed
- 20 asynchronously with a user's access to said hyper-text documents.
6. A method as claimed in claim 1, wherein said steps are performed substantially in synchronism with a user's access to said hyper-text documents.
- 25 7. A method as claimed in claim 1, wherein said formatted version is formatted to be suitable for single or multiple column page printing on a printer output device.

8. A method as claimed in claim 7, wherein said formatted version suitable for single or multiple column page printing comprises as many hyper-text documents on each printed page as can reasonably fit in a space available on said each printed page.

5 9. A method as claimed in claims 1, wherein said formatted version includes a table of contents listing each hyper-text document represented in said formatted version wherein each entry in the said table of contents is labelled with the position at which the associated hyper-text document occurs within the said formatted version.

10 10. A method as claimed in claim 1, wherein said formatted version includes a hyper-link index of all the hyper-link references in all the said hyper-text documents represented in said formatted version.

11. A method as claimed in claim 10, wherein each hyper-link reference in the said
15 formatted version is tagged with a cross-reference to its entry in said hyper-link index.

12. A method as claimed in claim 10, wherein said hyper-link index excludes hyper-link references of hyper-text documents represented in said formatted version.

20 13. A method as claimed in claim 1, wherein the said hyper-text documents are HTML documents.

14. A method as claimed in claim 1, wherein the said hyper-text documents are accessed using Internet protocols.

25

15. A method as claimed in claims 1, wherein said formatted version is displayed in preview form continuously while said user accesses said hyper-text documents.

16. A method of collating hyper-text documents, said method comprising steps of:

(a) accessing said hyper-text documents including structure information;

(b) creating a formatted version of said accessed hyper-text documents wherein said formatted version is characterised by a single or multiple column printing such that each printed page contains as many of said hyper-text documents as can reasonably fit
5 in an available space on a printed page.

17. A method as claimed in claim 16, wherein said hyper-text documents are determined by accepting a specification from a user of one or more root hyper-text documents and adding to said root hyper-text documents all derived hyper-text
10 documents which are hyper-linked from said root hyper-text documents and have certain specified characteristics defined by said user.

18. A method as claimed in claims 16, wherein said formatted version includes a table of contents listing each hyper-text document represented in said formatted version
15 wherein each entry in the said table of contents is labelled with the position at which the associated hyper-text document occurs within the said formatted version.

19. A method as claimed in claim 16, wherein said formatted version includes a hyper-link index of all the hyper-link references in all the said hyper-text documents
20 represented in said formatted version.

20. A method as claimed in claim 16 wherein each hyper-link reference in the said formatted version is tagged with a cross-reference to its entry in said hyper-link index.

21. A method as claimed in claim 16, wherein said hyper-link index excludes hyper-
25 link references of hyper-text documents represented in said formatted version.

22. A method as claimed in claim 16, wherein the said hyper-text documents are HTML documents.

23. A method as claimed in claim 16, wherein the said hyper-text documents are accessed using Internet protocols.

5 24. A method as claimed in claims 16, wherein said formatted version is displayed in preview form continuously while said user accesses said hyper-text documents.

25. Apparatus configured to implement the method of claim 1.

10 26. Apparatus configured to implement the method of claim 16.

27. A computer implemented method for collating a plurality of documents obtained from a plurality of sources, said method comprising the steps of:

monitoring accesses to documents in sequence;

15 recording the contents of a plurality of selected documents including structure information relating to each said selected document; and

collating said selected documents according to a predetermined order of collation, said collating comprising arranging none or more display pages according to the sizes of each said selected document based upon said corresponding structure information, 20 wherein said collating forms a single document reproducible at least by printing.

28. A computer system comprising:

a network comprising a source of a plurality of documents each individually accessible via a resource locator, wherein ones of said documents include therein links 25 that give access to others of said documents;

means for monitoring said resource locator and compiling a display list of said documents, said list including the corresponding links and structure information pertaining to each document; and

means for collating the display list into a selected order and for formatting said documents within said display list into a single printable document having corresponding components arranged in said selected order.

5 30. A computer readable medium including instruction modules arranged to collate for printing as a single document a plurality of documents derived from a plurality of sources in a network, said modules comprising:

a monitoring module for monitoring browsing operations throughout said network;

10 a compiling module for compiling a display list of selected documents encountered during said browsing operations; and

a collating module for collating the selected documents into a single printable document in which each said selected document is formatted according to structure information derived during said monitoring module whereby said single printable
15 document is collated to be substantially seamless in printing reproduction and to minimize vacant or wasted space on any and each printed page.

31. A medium as claimed in claim 30 wherein said medium is one of a computer network, a hard disk, a floppy disk and an optical disk.

20

32. A computer program product having a computer readable medium having a computer program recorded thereon for collating hyper-text documents, said computer program product comprising:

means for monitoring a user's access patterns to said hyper-text documents;

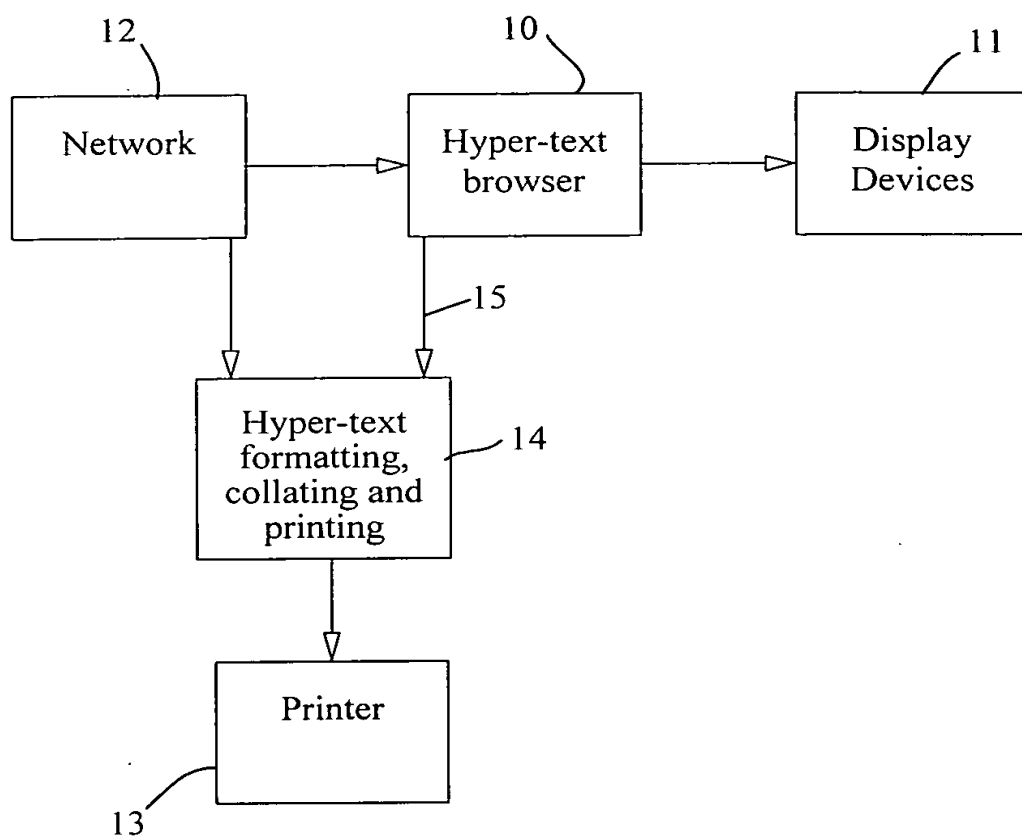
25 means for accessing said hyper-text documents including structure information of the accessed hyper-text documents; and

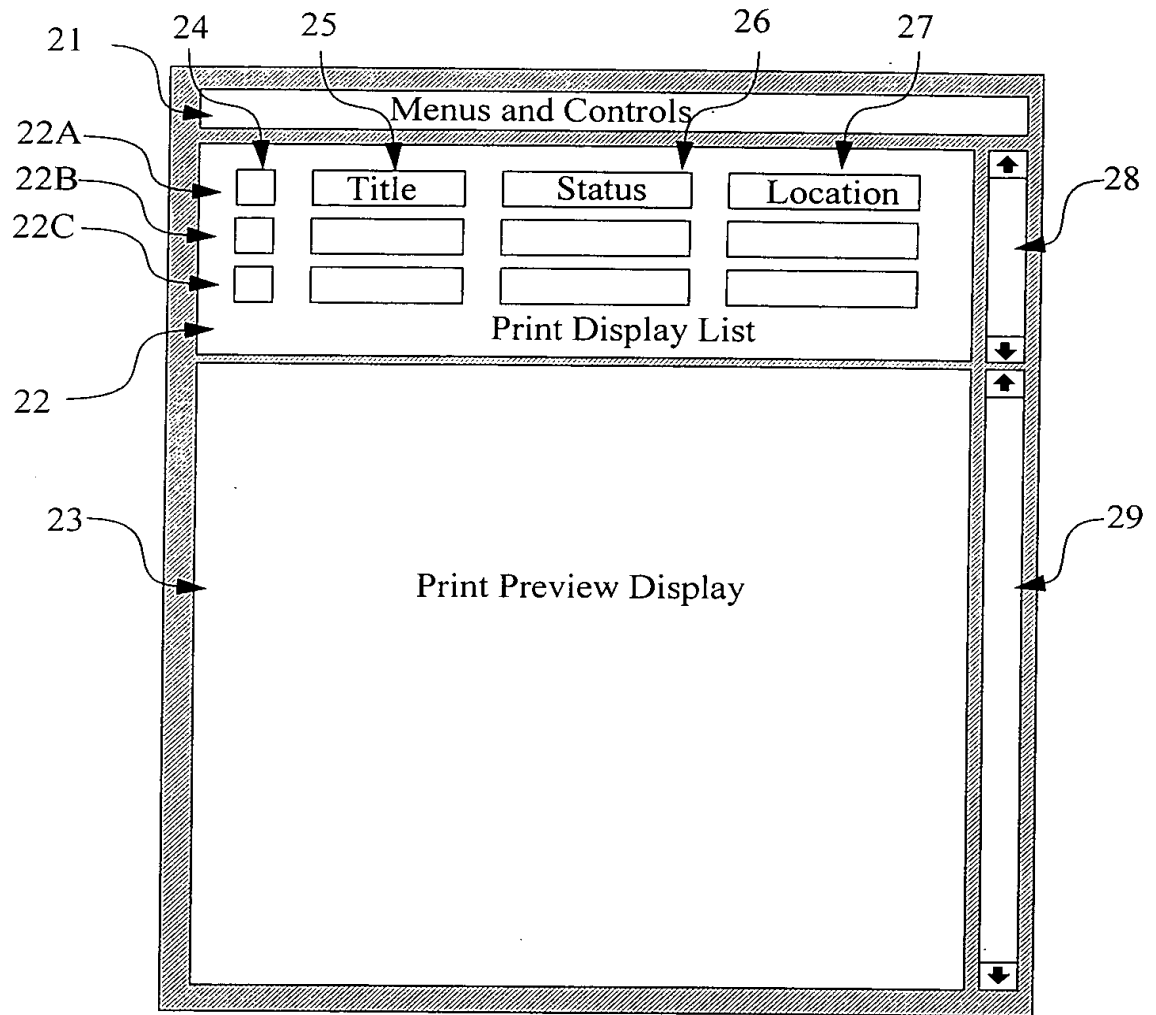
means for creating a formatted version of the accessed hyper-text documents for said user.

Abstract

HYPER-TEXT DOCUMENT FORMATTING COLLATING AND PRINTING

Disclosed is a method and apparatus for formatting, collating and printing, on an
5 output print device, hyper-text documents in a format favouring a printed document.
The method includes: (a) monitoring a user's access patterns to said hyper-text
documents; (b) accessing said hyper-text documents including structure information of
the accessed hyper-text documents; (c) creating a formatted version of the accessed
hyper-text documents for said user. Preferably the documents are "harvested", or
10 fetched from various hyper-links, in a background mode while a user is accessing
various hyper-text documents.

**FIG. 1**

**FIG. 2**

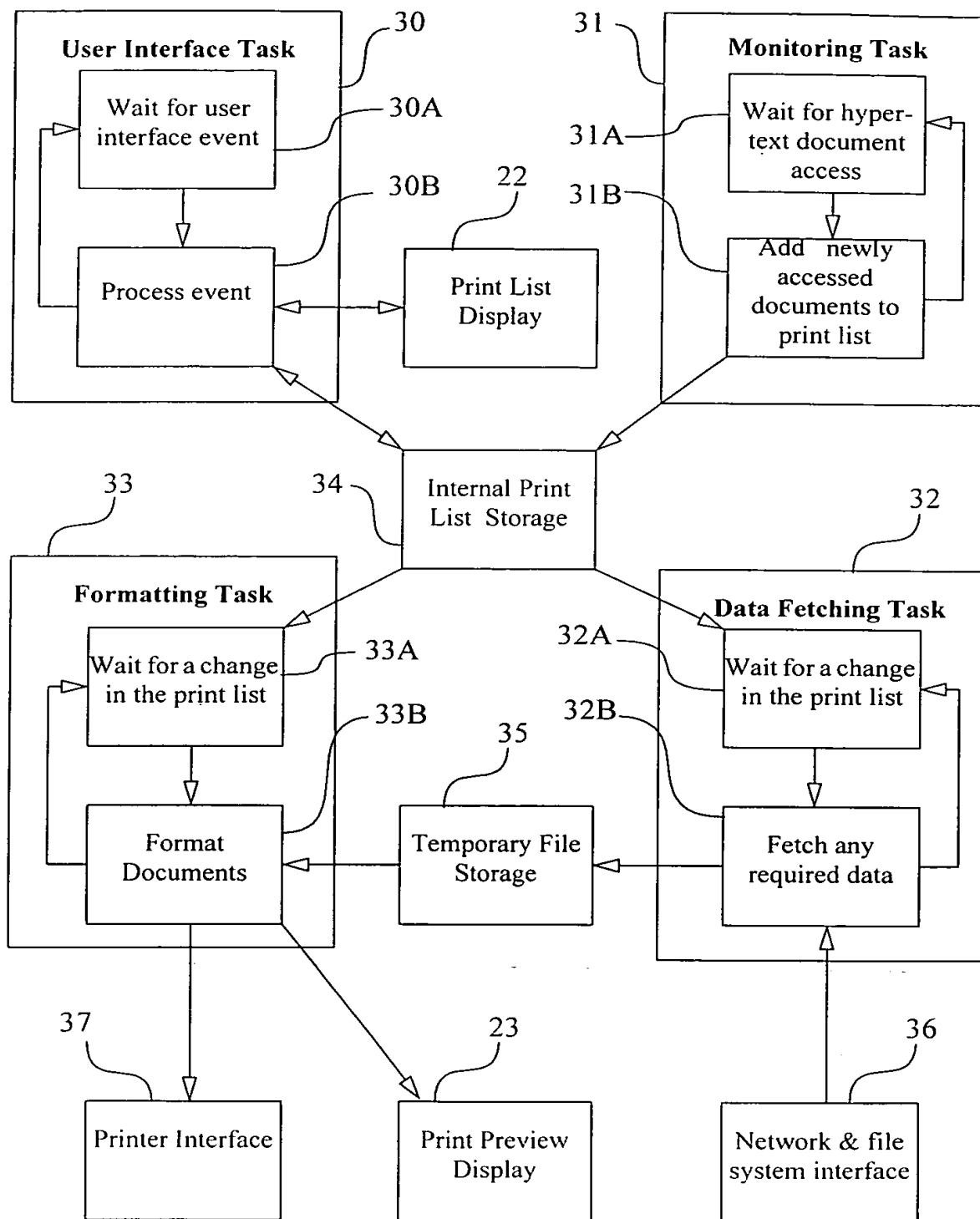
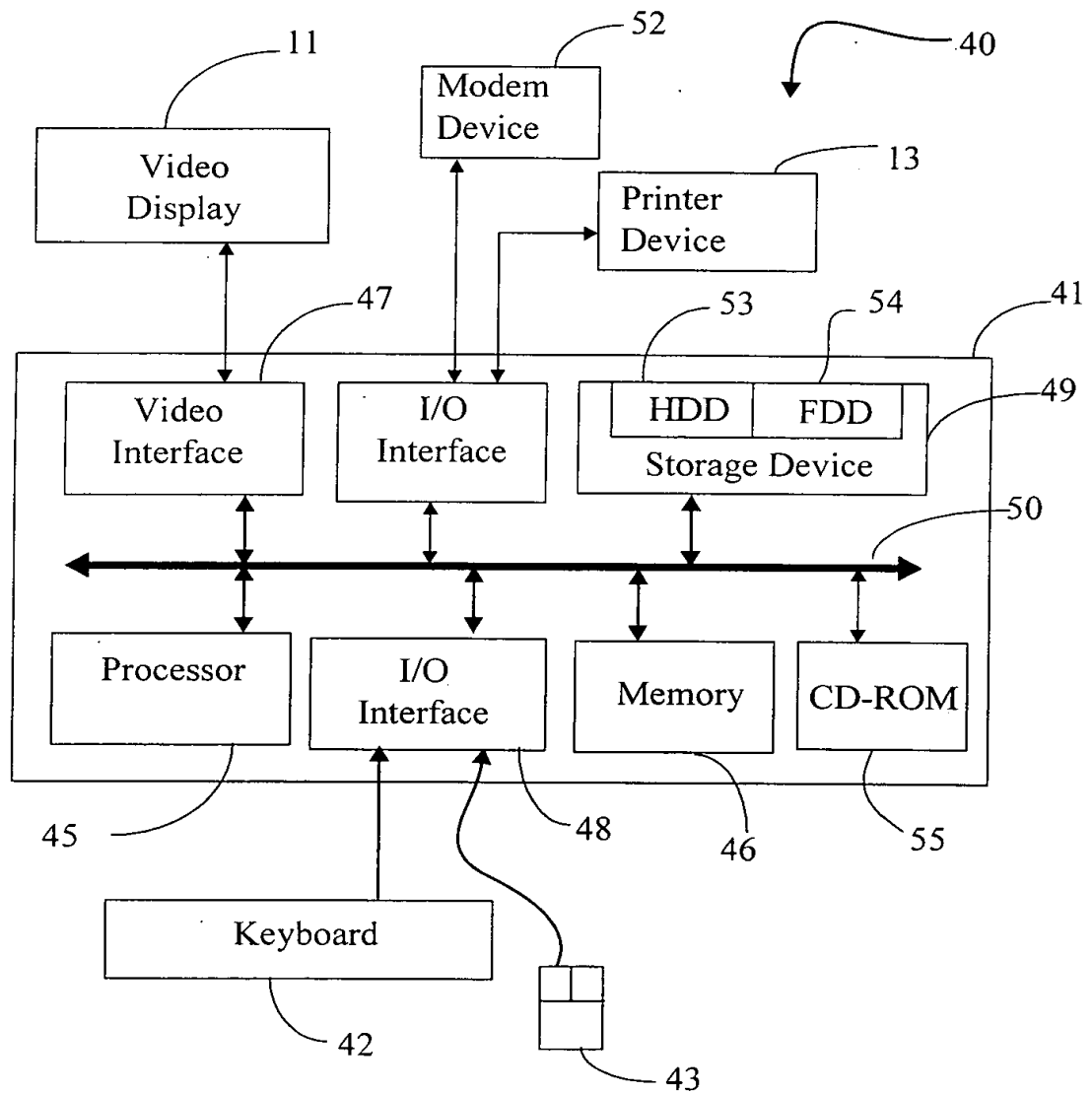


FIG. 3

**FIG. 4**

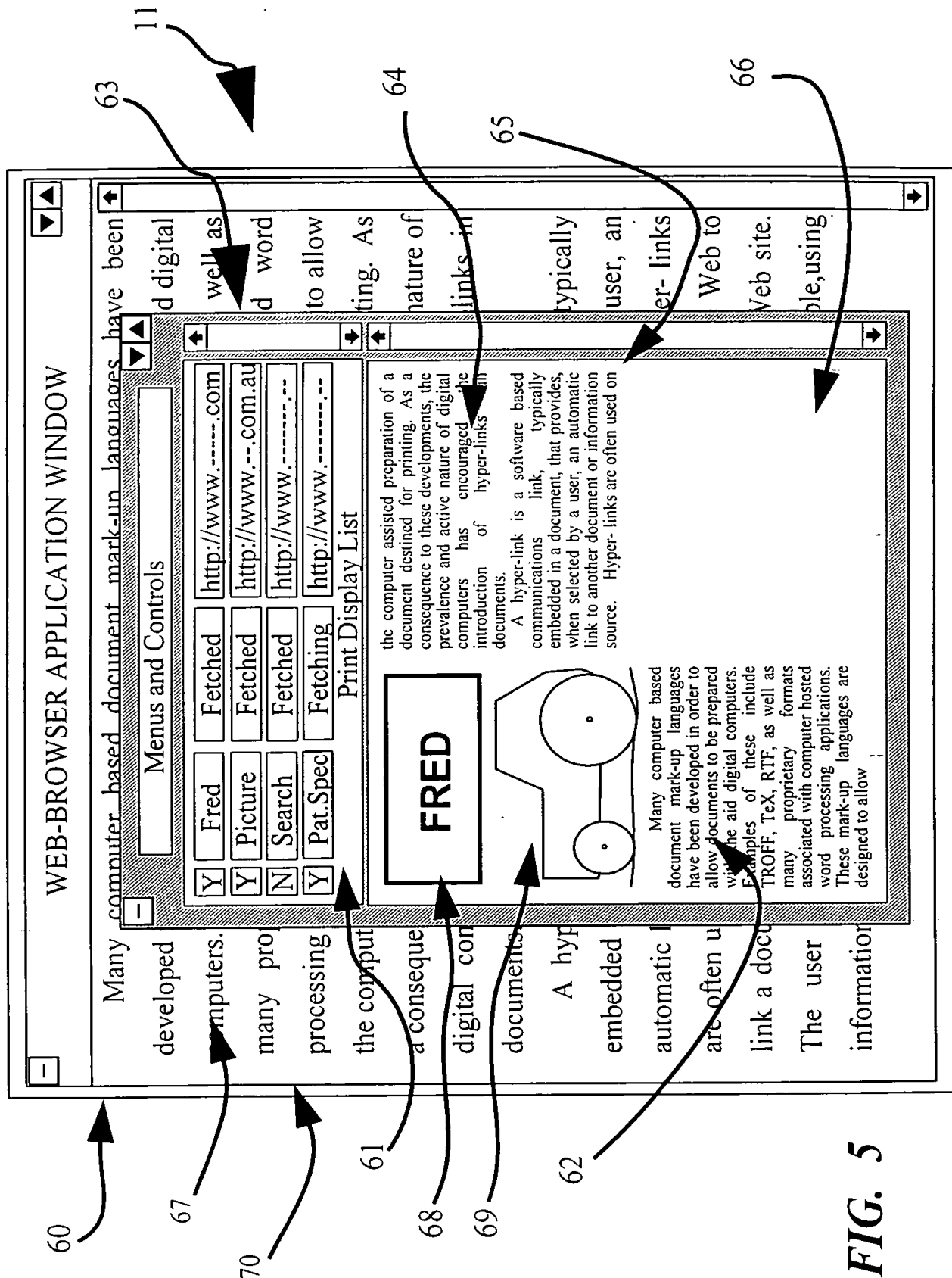
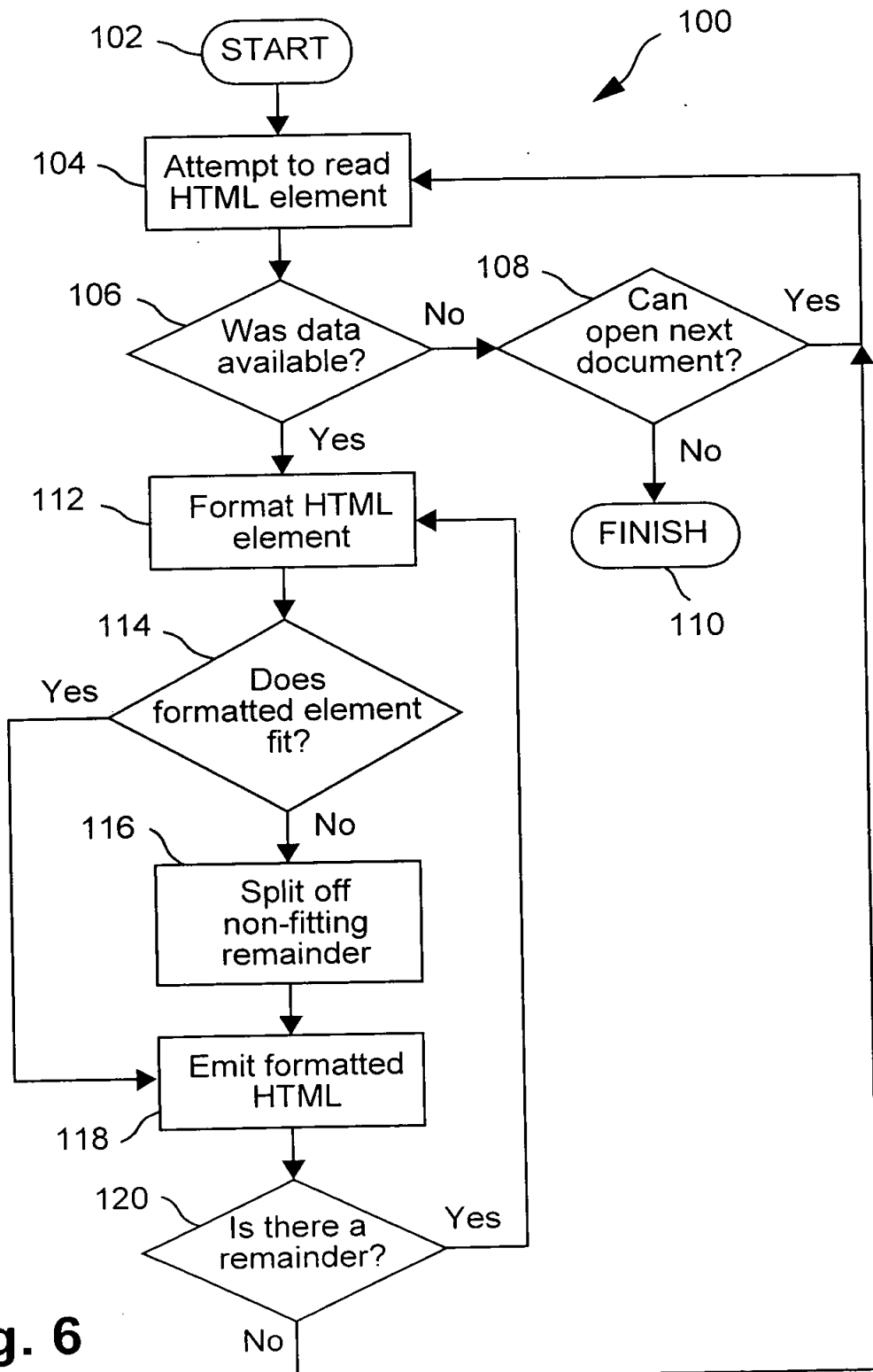


FIG. 5

**Fig. 6**

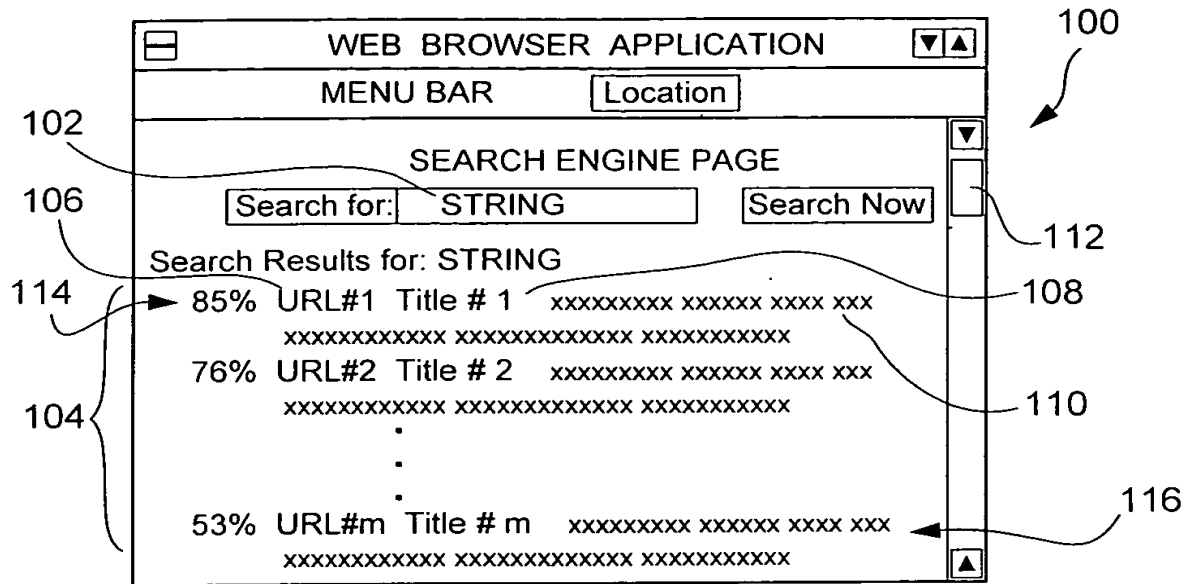


FIG. 1 (Prior Art)

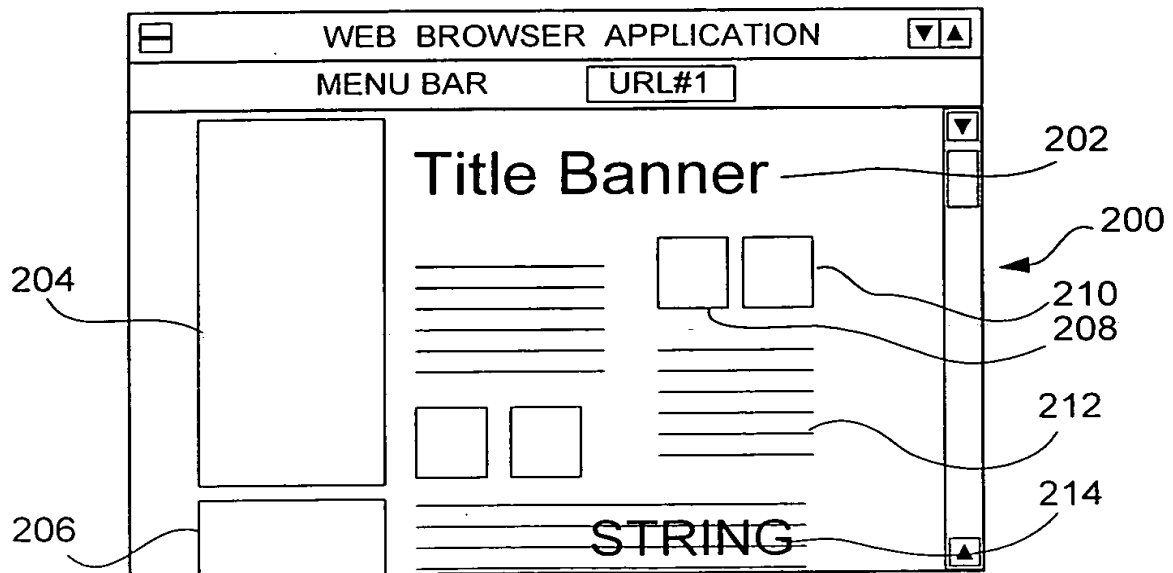


FIG. 2A (Prior Art)

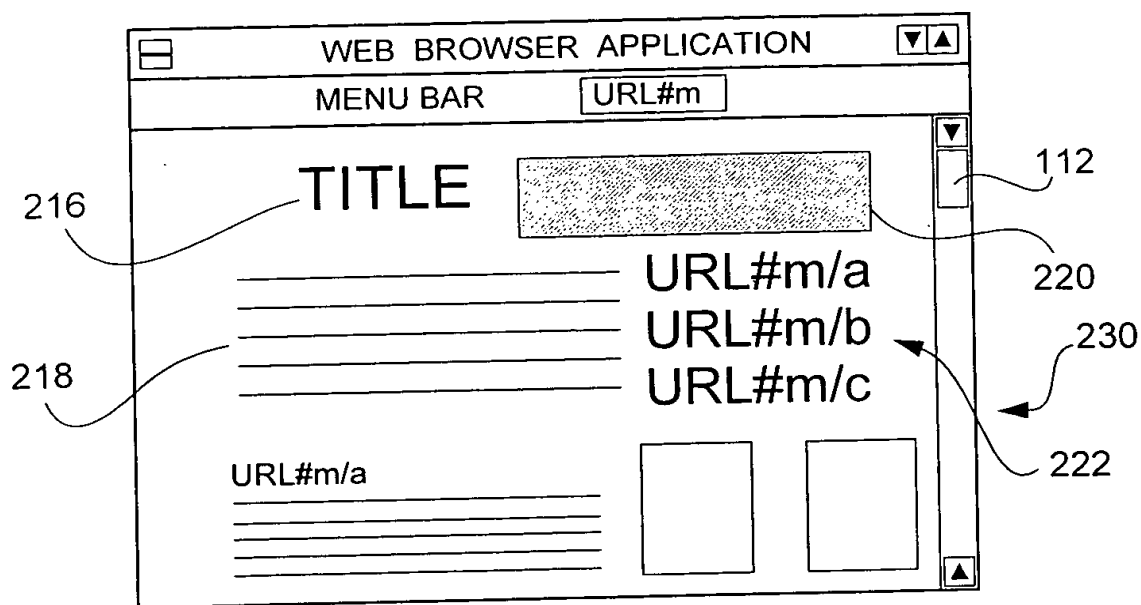


FIG. 2B (Prior Art)

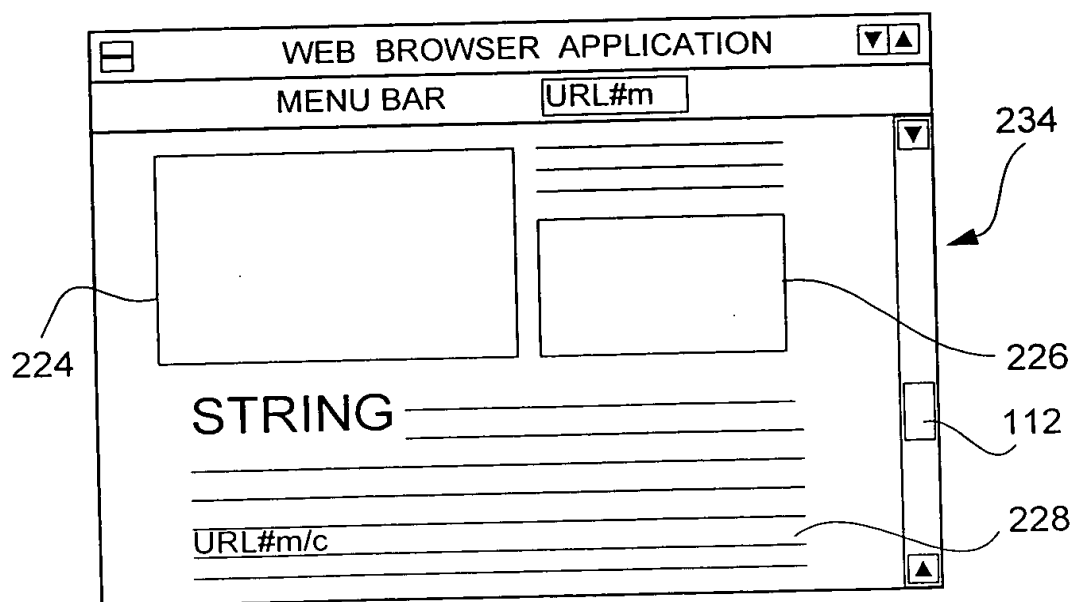


FIG. 2C (Prior Art)

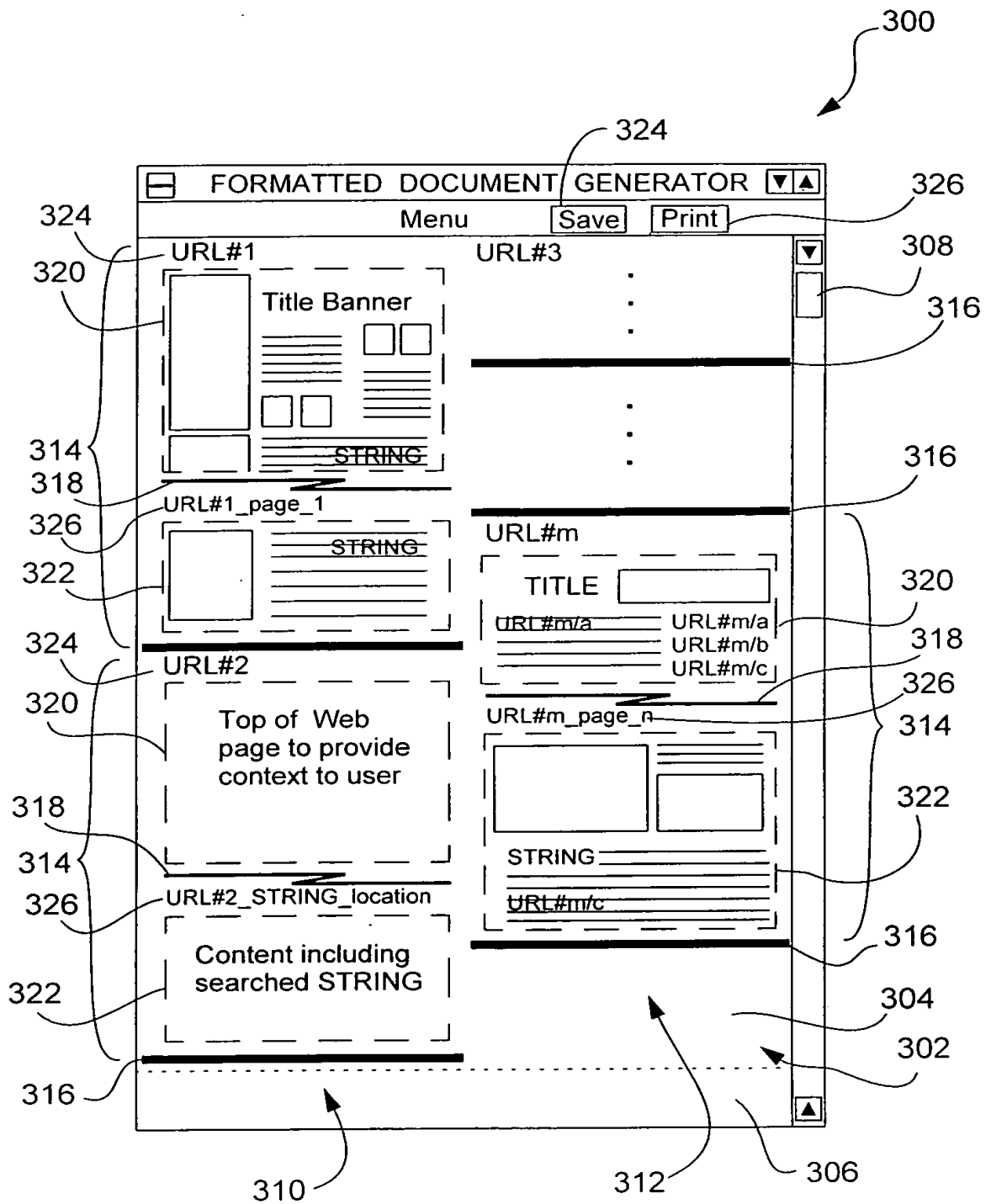
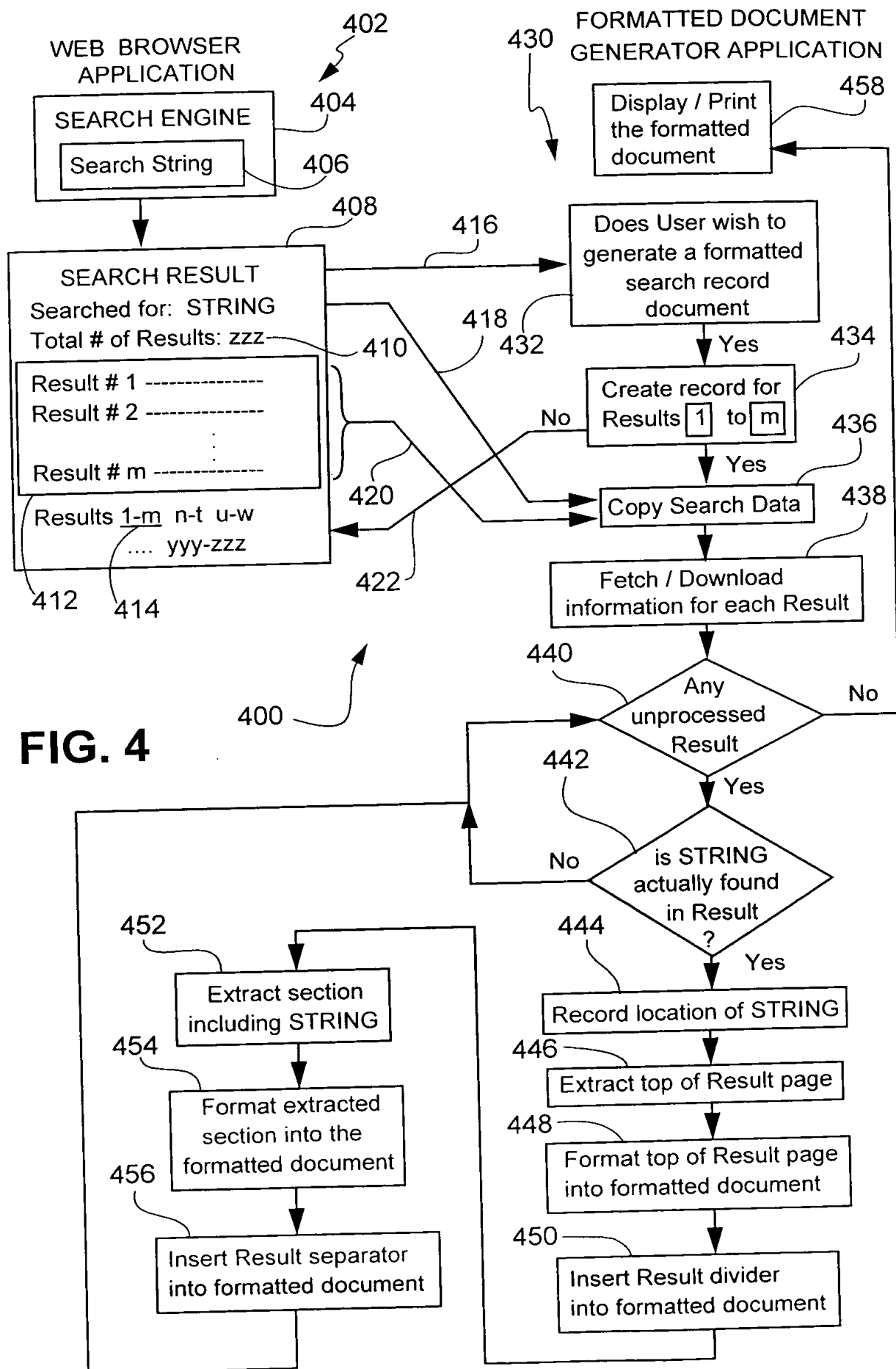


FIG. 3



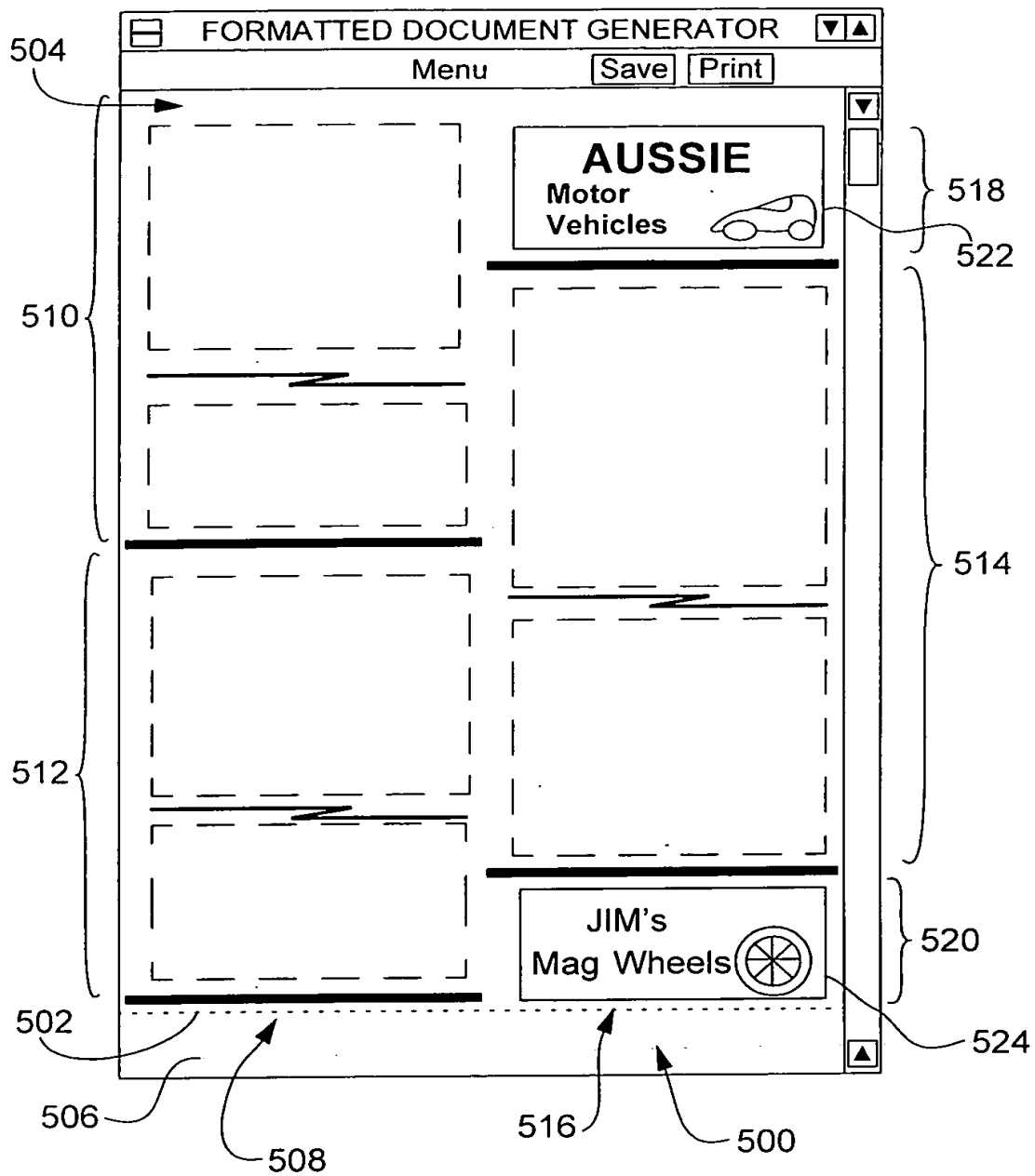
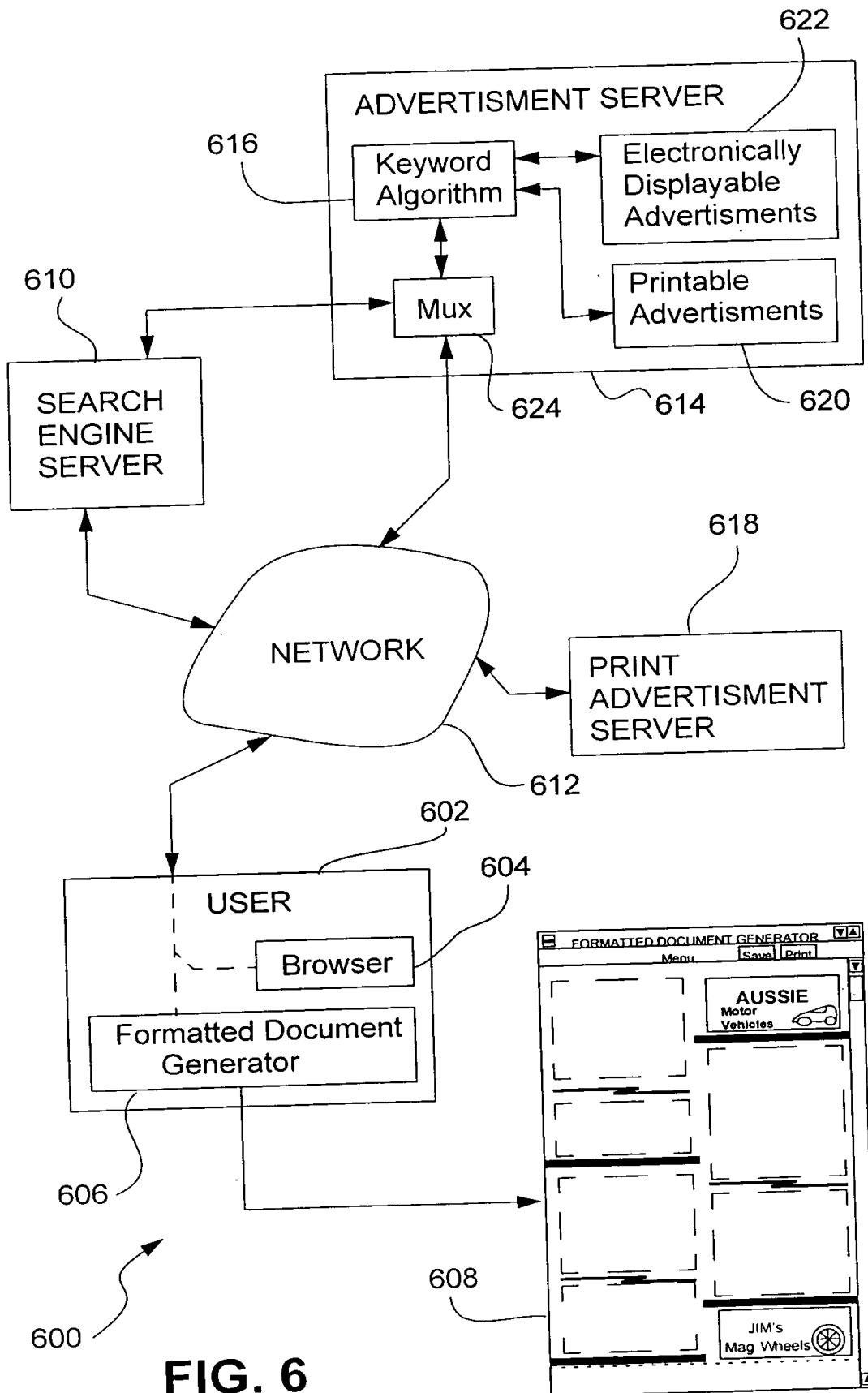


FIG. 5

**FIG. 6**

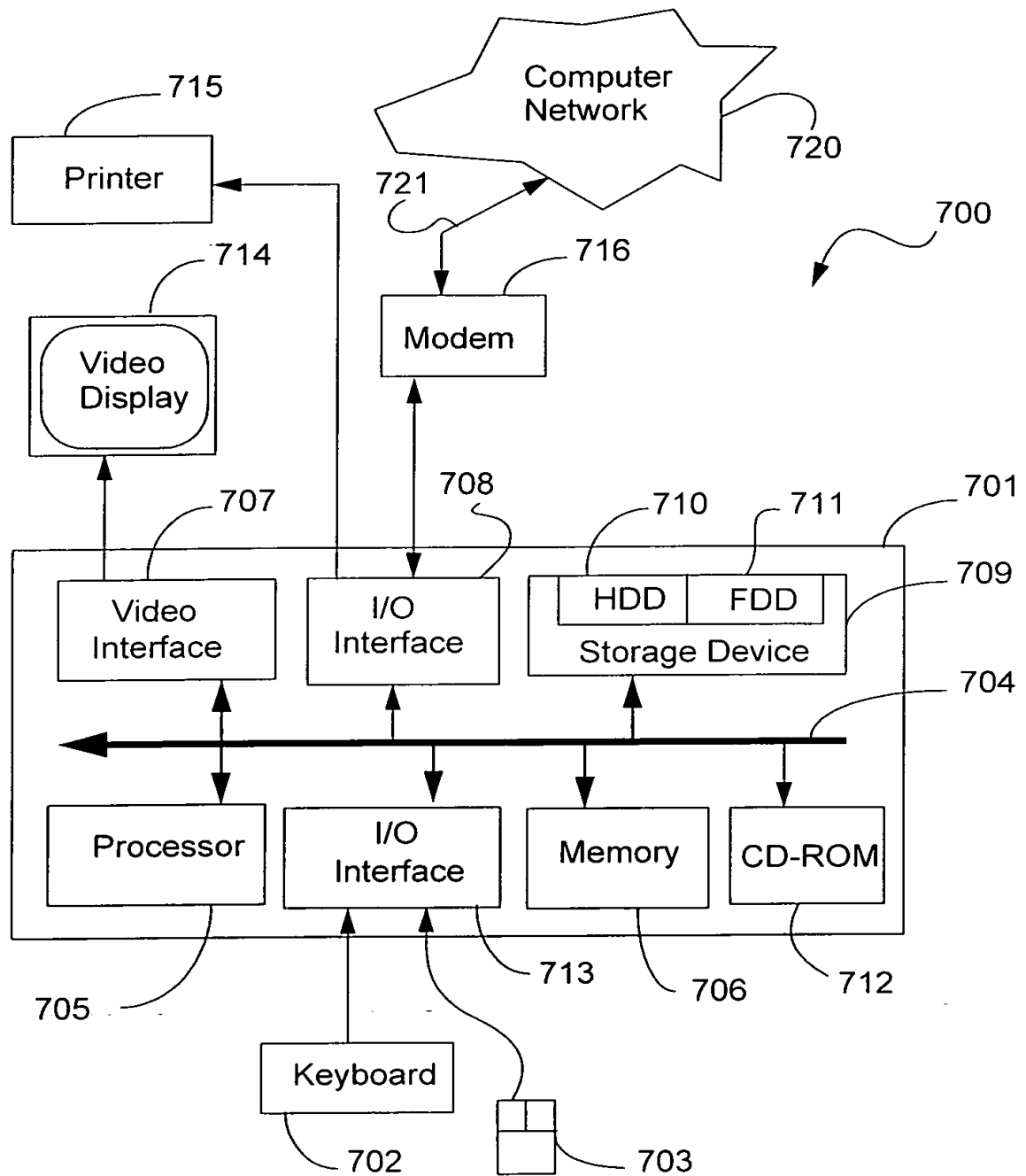


FIG. 7

This Page Blank (uspto)